



Success of *Escherichia coli* O25b:H4 Sequence Type 131 Clade C Associated with a Decrease in Virulence

Marion Duprilot,^{a,b} Alexandra Baron,^a François Blanquart,^{a,c} Sara Dion,^a Cassandra Pouget,^d Philippe Lettéron,^e Saskia-Camille Flament-Simon,^f Olivier Clermont,^a  Erick Denamur,^{a,g}  Marie-Hélène Nicolas-Chanoine^a

^aUniversité de Paris, INSERM, IAME, Paris, France

^bAP-HP, Laboratoire de Microbiologie, Hôpital Beaujon, Clichy, France

^cCentre for Interdisciplinary Research in Biology (CIRB), Collège de France, CNRS, INSERM, PSL Research University, Paris, France

^dVBMI, INSERM U1047, Université de Montpellier, Nîmes, France

^eUniversité de Paris, UMR 1149, INSERM-ERL, CNRS 8252, Paris, France

^fLaboratorio de Referencia de *Escherichia coli* (LREC), Departamento de Microbiología e Parasitología, Facultad de Veterinaria, Universidade de Santiago de Compostela (USC), Lugo, Spain

^gAP-HP, Laboratoire de Génétique Moléculaire, Hôpital Bichat, Paris, France

Marion Duprilot and Alexandra Baron contributed equally to this work. The order of the authors was determined by the greatest contribution of each of them in relation to the number of years required to carry out this work.

ABSTRACT *Escherichia coli* O25b:H4 sequence type 131 (ST131), which is resistant to fluoroquinolones and which is a producer of CTX-M-15, is globally one of the major extraintestinal pathogenic *E. coli* (ExPEC) lineages. Phylogenetic analyses showed that multidrug-resistant ST131 strains belong to clade C, which recently emerged from clade B by stepwise evolution. It has been hypothesized that features other than multidrug resistance could contribute to this dissemination since other major global ExPEC lineages (ST73 and ST95) are mostly antibiotic susceptible. To test this hypothesis, we compared early biofilm production, presence of ExPEC virulence factors (VFs), and *in vivo* virulence in a mouse sepsis model in 19 and 20 epidemiologically relevant strains of clades B and C, respectively. Clade B strains were significantly earlier biofilm producers ($P < 0.001$), carriers of more VFs ($P = 4e-07$), and faster killers of mice ($P = 2e-10$) than clade C strains. Gene inactivation experiments showed that the *H30-fimB* and *ibeART* genes were associated with *in vivo* virulence. Competition assays in sepsis, gut colonization, and urinary tract infection models between the most anciently diverged strain (B1 subclade), one C1 subclade strain, and a B4 subclade recombining strain harboring some clade C-specific genetic events showed that the B1 strain always outcompeted the C1 strain, whereas the B4 strain outcompeted the C1 strain, depending on the mouse niches. All these findings strongly suggest that clade C evolution includes a progressive loss of virulence involving multiple genes, possibly enhancing overall strain fitness by avoiding severe infections, even if it comes at the cost of a lower colonization ability.

KEYWORDS *Escherichia coli*, ST131, fimbriae, *ibeART* operon, *in vivo* virulence, niche adaptation

In the last 2 decades, multidrug-resistant (MDR) *Escherichia coli* O25b:H4 sequence type 131 (ST131) has emerged worldwide among human extraintestinal pathogenic *E. coli* (ExPEC) strains (1, 2). This *E. coli* lineage is composed of two clades, called clades B and C, defined on the basis of the *fimH* alleles H22 and H30 (3, 4). A time-calibrated phylogeny showed that clade C emerged from clade B and that sequential mutational events shaped the evolution of O25b:H4 ST131 from the 1950s, resulting in the diversification of clade B into different subclades (subclades B1 to B5 and then the

Citation Duprilot M, Baron A, Blanquart F, Dion S, Pouget C, Lettéron P, Flament-Simon S-C, Clermont O, Denamur E, Nicolas-Chanoine M-H. 2020. Success of *Escherichia coli* O25b:H4 sequence type 131 clade C associated with a decrease in virulence. *Infect Immun* 88:e00576-20. <https://doi.org/10.1128/IAI.00576-20>.

Editor Andreas J. Bäumler, University of California, Davis

Copyright © 2020 Duprilot et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Marie-Hélène Nicolas-Chanoine, marie-helene.nicolas-chanoine@inserm.fr.

Received 14 September 2020

Accepted 14 September 2020

Accepted manuscript posted online 28 September 2020

Published 16 November 2020

intermediate subclade, B0) and the appearance of clade C (5). The diversification of clade C into different subclades was characterized by the acquisition of the *fimH30* allele in about 1980, resulting in the C0 subclade, followed by the acquisition of the *gyrA1AB* and *parC1aAB* alleles, encoding fluoroquinolone resistance, in about 1987, resulting in the C1 and C2 subclades (5). All C2 subclade strains and a fraction of the C1 subclade strains forming the C1-M27 cluster (6) are additionally resistant to extended-spectrum cephalosporins (ESC) due to the production of extended-spectrum β -lactamases (ESBL) of the CTX-M type, CTX-M-15 and CTX-M-27, respectively.

Existing epidemiological studies have examined the prevalence of ST131 among *E. coli* isolates that are resistant to fluoroquinolones and/or that are producers of ESBL (7). Few studies have assessed the relative frequencies of ST131 clade B and clade C (8–11), and none of them have assessed the relative frequencies of the B and C subclades.

The evolutionary success of MDR ST131 clade C is still largely unexplained. Multidrug resistance may explain this success, as resistance to fluoroquinolones and ESC was shown to not impact the growth fitness of clade C strains (12, 13). However, *E. coli* lineages susceptible to antibiotics, such as ST73 and ST95, are as successful as ST131 clade C (in terms of their frequency in bacteremia and urinary tract infections), suggesting that factors other than antibiotic resistance can participate in the success of a lineage (9). Thus, by combining population genomics and modeling approaches, McNally et al. (14) showed that, compared to other dominant *E. coli* lineages and also to ST131 clade B, MDR ST131 clade C has accumulated a significantly elevated allelic diversity particularly enriched for genes involved in anaerobic metabolisms and other loci encoding factors contributing to human colonization by ExPEC. Given the low frequency of occurrence of this allelic diversity, the authors suggested adaptation to separate ecological niches among ST131 clades. For their part, Billard-Pomares et al. (15) showed that since 2000 the arginine deiminase operon has been increasing in prevalence in isolates producing ESBL, especially CTX-M, largely due to its presence in ST131 clade C. This operon was deleterious for the ESBL-producing strain tested in a mouse gut colonization model but contributed to its virulence in a mouse model of urinary tract infection (15). Therefore, this suggests the potential for ST131 clade C to adapt to different human niches.

In the present study, we chose to mostly perform phenotypic comparisons between clade B and clade C strains to get further insights into factors that have favored clade C dissemination. To this end, we established a collection of 39 *E. coli* O25b:H4 ST131 strains composed of strains representing the O25b:H4 ST131 clade and subclade diversity. The characterization of the 39 strains with regard to early biofilm formation and the capability to kill mice in a sepsis model, as well as competitions between selected clade B and clade C strains in various mouse models, allowed us to highlight the difference in the virulence potential of clade B and clade C strains.

RESULTS

Epidemiology of the O25b:H4 ST131 B and C subclades in extraintestinal infections over time. Recently, Kallonen et al. depicted 197 O25b:H4 ST131 isolates (161 belonging to clade C and 36 belonging to clade B) among 1,509 bacteremia *E. coli* isolates systematically collected between 2001 and 2012 in England (9). To have an overview of the epidemiology of the O25b:H4 ST131 B and C subclades, we used the whole-genome sequence of the English O25b:H4 ST131 isolates and reference strains of the different O25b:H4 ST131 B and C subclades (5) to build a phylogenetic tree from nonrecombinant single nucleotide polymorphisms (SNPs) of core genomes and clarify to which B and C subclades the 197 English O25b:H4 ST131 isolates belonged (see Fig. S1 in the supplemental material). We then assessed their relative frequencies (Fig. 1A and B) and investigated the trends in their frequency over the period from 2001 to 2012 (Fig. 1C). Among clade B strains, only the B5 subclade strains exhibited a notable increase (logistic regression for B5 strain presence as a function of time, 0.24 per year [95% confidence interval {CI}, –0.005 to 0.5 per year]). Both the C1 and C2 subclade

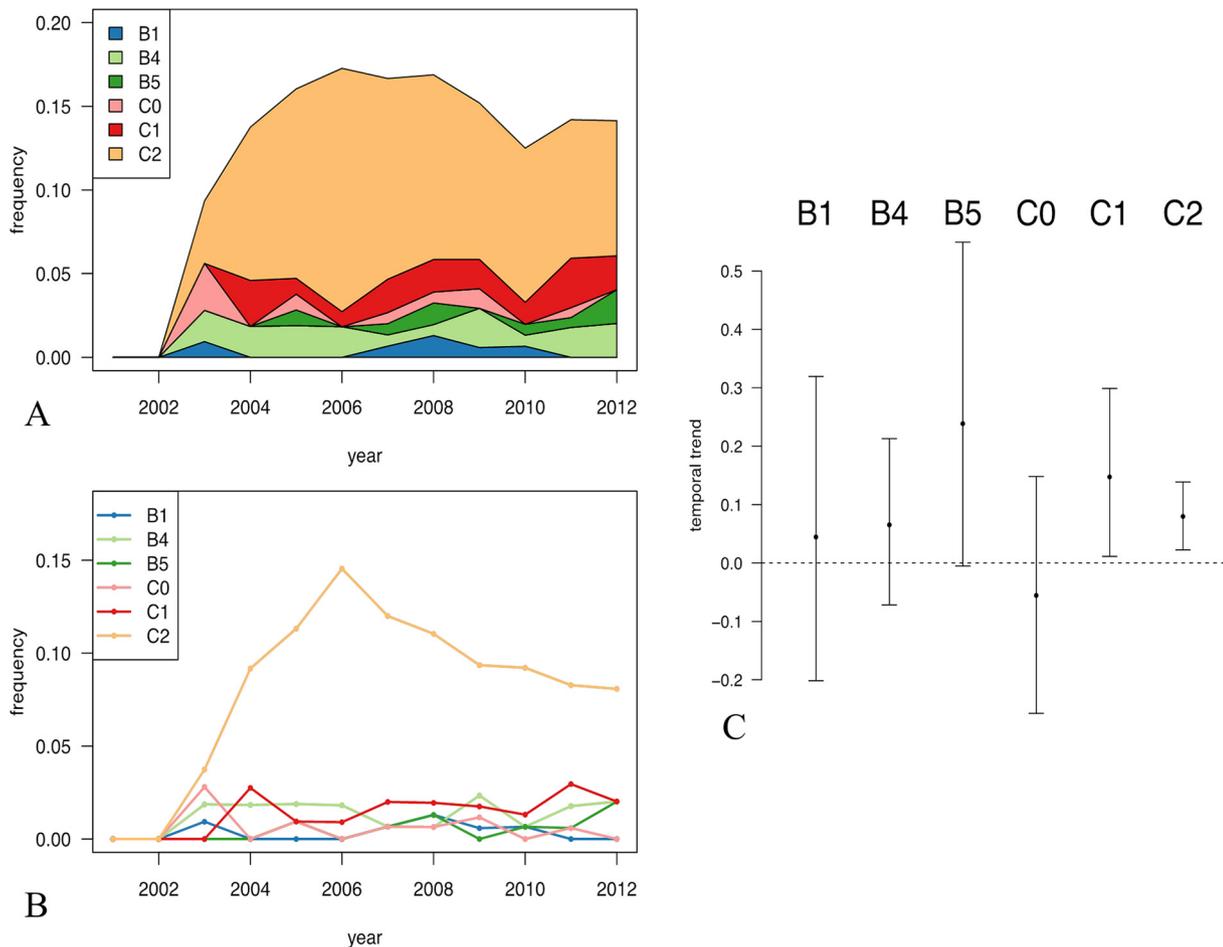


FIG 1 Dynamics of subclades in 197 O25b:H4 ST131 *Escherichia coli* strains isolated from cases of bacteremia over a 11-year sampling period (data are from Kallonen et al. [9]). (A) Cumulative frequencies of O25b ST131 subclades as a function of time, from the most ancestral B1 subclade at the bottom to the derived C2 subclade on the top. (B) Frequency of ST131 subclades as a function of time. Subclades are represented by colors. (C) Inferred linear trends of ST131 subclade frequencies (among all *E. coli* strains [$n = 1,509$]) as a function of time from 2001 to 2012. Among the clade B strains, only the subclade B5 strains exhibited a notable increase (logistic regression for subclade B5 strain presence as a function of time, 0.24 per year [95% CI, -0.005 to 0.55 per year]). Both the C1 and C2 subclade strains significantly increased (0.15 per year [95% CI, 0.01 to 0.30 per year] for C1, 0.08 per year [95% CI, 0.02 to 0.14 per year] for C2).

strains significantly increased (0.15 per year [95% CI, 0.01 to 0.30 per year] for C1, 0.08 per year [95% CI, 0.02 to 0.14 per year] for C2).

Molecular characteristics of the 39 studied O25b:H4 ST131 strains. As illustrated in Fig. 1 and by studies that had used the *fimH* typing method to characterize O25b:H4 ST131 collections (16, 17), clade B/*fimH22* strains are much less frequent than clade C/*fimH30* strains. Therefore, to establish an O25b ST131 collection comprising as many clade B strains as clade C strains, we obtained *fimH22* strains either previously published or belonging to the coauthors' personal collections (Table S1). Thus, we were able to get 18 *fimH22* strains, to which were added 21 *fimH30* strains, which generally came from the same collections as the *fimH22* strains (Table S1). The 39 strains were human strains isolated in Spain and France between 1993 and 2012 mostly from the blood of adult patients and the feces of healthy adults and children (Table S1). Through the phylogenetic tree constructed from nonrecombinant SNPs of core genome genes (Fig. S1), we found that the 39 strains were composed of 19 clade B strains distributed into the different B subclades, except for the B0 and B2 subclades (two subclades also absent among the 36 English O25b:H4 ST131 clade B isolates), and 20 clade C strains distributed into the three C subclades (subclades C0, C1, and C2). Among the 19 clade B strains, 17 displayed the *fimH22* allele; 1 displayed a *fimH22*-like allele; and 1, B4

subclade strain CES131C, displayed the *fimH30* allele that had been described to be specific to clade C strains (Table 1). For this reason, the CES131C strain is called "Recombinant" in the rest of this report.

All strains but one (strain 208) harbored an IncF plasmid with an FII, FIA, and FIB replicon allele composition corresponding to the one previously described for clade B and C0/C1 and C2 subclade strains (13), except for one C1 subclade strain (CES9C) that lacked the FIB replicon observed in C2 subclade strains (Table 1). Four clade B strains (two of the B1 subclade, one of the B3 subclade, and one of the B5 subclade) harbored a RepF1C replicon that belonged to the RepFIIA family (18), and one B3 subclade strain harbored two replicons of the RepFIIA family (Table 1).

As shown in Table 1, the 39 strains comprised 30 strains, 10 of which were clade B, displaying diverse antibiotic resistance-encoding genes with a gradual accumulation from clade B to clade C.

***In vitro* phenotypes of the 39 studied O25b:H4 ST131 strains.** We first assessed four *in vitro* phenotypes of the 39 collected strains: the maximum growth rate (MGR), the kinetics of early biofilm formation, and the expression of both type 1 and curli fimbriae.

For the MGR measured in a complex medium (lysogeny broth [LB]) under planktonic and shaking conditions (Fig. 2), there was no significant difference regardless of the comparisons made: clade B versus clade C ($P = 0.08$) (Fig. 2A), clinical versus fecal strains ($P = 0.07$) (Fig. 2B), nalidixic acid-resistant versus nalidixic acid-susceptible clade B strains ($P = 0.9$) (Fig. 2C), and nalidixic acid-/ciprofloxacin-resistant versus nalidixic acid-/ciprofloxacin-susceptible clade C strains ($P = 0.9$) (Fig. 2D).

We investigated the kinetics of early biofilm formation because previous results had shown that among a large collection of *E. coli* clinical isolates composed of 99 sequence type (ST) lineages, the O25b:H4 ST131 lineage belongs to the few lineages able to form a biofilm early in the experiment (19). We showed here that clade B strains were more frequently early biofilm producers than clade C strains ($P < 0.001$). More specifically, we found three significantly different phenotypes of early biofilm production ($P < 0.0001$) among the 39 studied strains (Fig. 3A; Table 1): early and persistent producers (biofilm index [BFI] ≤ 10 at 2, 3, and 5 h), delayed producers (BFI > 10 at 2 and 3 h but BFI ≤ 10 at 5 h), and never producers (BFI > 10 at 2, 3, and 5 h). As BFI values after 2 and 3 h of incubation classified a given strain into the same early biofilm formation phenotype (Fig. 3A), only BFI values obtained after 2 and 5 h of incubation were considered for further analyses. The 19 clade B strains were either early producers ($n = 8$), delayed producers ($n = 8$), or never producers ($n = 3$), whereas the 20 clade C strains were either delayed producers ($n = 8$) or never producers ($n = 12$) (Fig. 3A; Table 1). Within clade B, there was a significant difference between the seven B5 subclade strains (early producers, $n = 5$; delayed producers, $n = 2$) and the seven B4 subclade strains (delayed producers, $n = 4$; never producers, $n = 3$) ($P = 0.02$). Inversely, there was no significant difference between the three clade C subclades regarding the early biofilm formation phenotypes.

As the ability to form a biofilm is linked to type 1 and curli fimbria expression (20), we studied the expression of these types of fimbriae in the 39 strains at different times and under different growth conditions. After 2 and 5 h of shaking growth, the early biofilm producers more frequently expressed type 1 fimbriae than the delayed ($P < 0.01$) and never ($P \leq 0.0001$) producers, and clade B strains more frequently expressed type 1 fimbriae than clade C strains ($P \leq 0.0001$) (Table 2). All these significant differences were still observed after 24 h of growth under shaking conditions, but no significant difference was observed under static conditions (Table 2). Curli fimbria expression (Table 3), as assessed by the Congo red morphotype, was significantly associated with the early biofilm formation phenotype ($P < 0.0001$). Significantly more clade B strains (47%) than clade C strains (15%) were positive by the Congo red assay ($P < 0.001$).

TABLE 1 Genotypic and phenotypic characteristics of the 39 O25b:H4 ST131 *Escherichia coli* strains^a

Strain	fimH operon type	β-Lactamase		Virotype	Allele	Gene(s) encoding resistance to:		Plasmid Inc system	pMLST	Col-like plasmid	VF score	Biofilm phenotype
		Non-ESBL	ESBL			Aminoglycosides	Other antibiotics					
B11J12	H22	TEM-1B		D3	<i>gyrA1A</i>	<i>parC1</i>	<i>aac(3)-IIVa, aph(4)-Ia</i>	F1C, FIB	F89A-:B62	156	16	- +
S250	H22			D3	<i>gyrA1a</i>	<i>parC1</i>		F1C, FIB	F89A-:B62		14	- +
CE5106C2	H22		CTX-M-1	D4	<i>gyrA1a</i>	<i>parC2</i>	<i>sul2, tet(A)</i>	F1C, FIB, I1	F18A-:B1, I1-:ST3	pVC	16	- +
H2381	H22			D4	<i>gyrA1A</i>	<i>parC2</i>	<i>tet(A)</i>	FII, FIB, X1		MG828	18	- +
H3345	H22			D4	<i>gyrA1A</i>	<i>parC2</i>		FII, FIB, FIC	F18A-:B1	pVC	19	- +
H1659	H22		SHV-12	D3	<i>gyrA1a</i>	<i>parC1</i>		FII, FIB	F29A-:B10	156	16	- +
B7J19	H22			D3	<i>gyrA1a</i>	<i>parC1</i>		FII, FIB	F29A-:B10	156	16	-
B12C27	H22			D3	<i>gyrA1a</i>	<i>parC1</i>		FII, FIB	F29A-:B10	156, MG828	16	-
B5E11	H22			D3	<i>gyrA1a</i>	<i>parC1</i>		FII, FIB	F29A-:B10	156	16	-
011-005	H22			D3	<i>gyrA1a</i>	<i>parC1</i>		FII, FIB	F29A-:B10	156	15	- +
001-001	H22-like			ND	<i>gyrA1a</i>	<i>parC1</i>		FII, FIB	F29A-:B10	156	15	- +
CE5131C	H30			B4	<i>gyrA1a</i>	<i>parC1</i>	<i>aac(3)-IId, aadA2</i>	FII, FIB	F2A-:B10	BS512	15	- +
H219B	H22		TEM-1B	D1	<i>gyrA1a</i>	<i>parC1</i>	<i>mph(A), sul1, tet(A), dfrA12</i>	FII, FIB	F2A-:B10		15	- +
H1447	H22		TEM-1B	D1	<i>gyrA1A</i>	<i>parC1</i>	<i>strA, strB, aadA5</i>	F1C, FB1, Q1	F18A-:B8		17	- +
H1698	H22		TEM-1C	D2	<i>gyrA1a</i>	<i>parC1</i>	<i>tet(A)</i>	FII, FIB, X1	F2A-:B1		22	- +
196	H22		TEM-1B	D1	<i>gyrA1a</i>	<i>parC1</i>	<i>mph(A), sul1, tet(A), dfrA1</i>	FII, FIB	F2A-:B1		16	- +
208	H22			D1	<i>gyrA1a</i>	<i>parC1</i>		X1		156	12	- +
H2262	H22		CTX-M-9	D2	<i>gyrA1AB</i>	<i>parC1aAB</i>	<i>aac(3)-IId, aadA2</i>	FII, FIB, N, H12	F2A-:B1, N-:ST9, H12-:ST1	MG828	21	- +
005-019	H22			D1	<i>gyrA1a</i>	<i>parC1</i>	<i>qnrA1, sul1, tet(A), dfrA12</i>	FII, FIB	F24A-:B6		16	- +
H1088	H30			C2	<i>gyrA1a</i>	<i>parC1</i>	<i>aaaA5</i>	FII, FIB	F36A1-:B20	156, BS512, IMG531	14	-
B1G9	H30			C2	<i>gyrA1a</i>	<i>parC1a</i>	<i>mph(A), sul1, dfrA17</i>	FII, FIB	F22A1-:B20	MG828	15	-
H2214	H30			C2	<i>gyrA1a</i>	<i>parC1a</i>	<i>aaaA5</i>	FII, FIB	F36A1-:B20	MG828	14	-
B1A5	H30			C3	<i>gyrA1AB</i>	<i>parC1aAB</i>	<i>aac(3)-IId</i>	FII, FIB	F2A-:B20	IMG531	12	- +
B1H12	H30			C3	<i>gyrA1AB</i>	<i>parC1aAB</i>	<i>aaaA5, strA, strB</i>	FII, FIB	F1:A2-:B20	156, IMG531	12	- +
B2B2	H30		CTX-M-14	C2	<i>gyrA1AB</i>	<i>parC1aAB</i>	<i>aaaA5</i>	FII, FIB	F1:A2-:B20, N-:UST	KPH56	14	-
02	H30			C3	<i>gyrA1AB</i>	<i>parC1aAB</i>	<i>aaaA5</i>	FII, FIB, N	F1:A2-:B20, N-:UST		12	- +
39	H30			C2	<i>gyrA1AB</i>	<i>parC1aAB</i>	<i>strA, strB, aadA5</i>	FII, FIB	F1:A2-:B20	BS512, KPH56	14	-
183	H30			C3	<i>gyrA1AB</i>	<i>parC1aAB</i>	<i>mph(A), sul1-sul2, tet(A), dfrA17</i>	FII, FIB	F1:A2-:B20	BS512, IMG531	13	-
187	H30			C3	<i>gyrA1AB</i>	<i>parC1aAB</i>	<i>mph(A), sul1-sul2, tet(A), dfrA17</i>	FII, FIB, I1, X1	F1:A2-:B20, I1-:ST196	BS512, IMG531	12	- +
CE5103C	H30			C3	<i>gyrA1AB</i>	<i>parC1aAB</i>	<i>strA, strB, aadA5</i>	FII, FIB, I1, X4	F1:A2-:B20, I1-:UST	BS512	12	-
CE59C	H30			C2	<i>gyrA1AB</i>	<i>parC1aAB</i>	<i>aac(3)-IId</i>	FII, FIB	F1:A2-:B20	IMG531	14	-
CE5164C	H30			C2	<i>gyrA1AB</i>	<i>parC1aAB</i>	<i>aac(3)-IId</i>	FII, FIB	F1:A2-:B20	BS512	13	-
H3084B	H30			C1	<i>gyrA1AB</i>	<i>parC1aAB</i>	<i>aac(3)-IId</i>	FII, FIB	F1:A2-:B20	BS512	11	- +
C5	H30		CTX-M-27	C2	<i>gyrA1AB</i>	<i>parC1aAB</i>	<i>strA, strB, aadA5</i>	FII, FIB	F1:A2-:B20, I1-:UST, N-:ST9	BS512	14	-
TN03	H30		CTX-M-15	C2	<i>gyrA1AB</i>	<i>parC1aAB</i>	<i>mph(A), sul1-sul2, tet dfrA17</i>	FII, FIB, I1, N	F1:A2-:B20, I1-:UST, N-:ST9	BS512	11	- +
B5B6	H30		CTX-M-15	C2	<i>gyrA1AB</i>	<i>parC1aAB</i>	<i>aac(3)-IId, aac(6')/Ib-cr</i>	FII, FIB, I1	F1:A2-:B20, I1-:ST193	BS512, IMG531	14	-
B1211	H30		CTX-M-15	C2	<i>gyrA1AB</i>	<i>parC1aAB</i>	<i>aac(3)-IId, aac(6')/Ib-cr</i>	FII, FIB	F2A1-:B-	BS512	14	-
C23	H30		CTX-M-15	C2	<i>gyrA1AB</i>	<i>parC1aAB</i>	<i>aac(6')/Ib-cr, aadA5</i>	FII, FIB	F2A1-:B-	IMG531	14	- +
C3	H30		CTX-M-15	C2	<i>gyrA1AB</i>	<i>parC1aAB</i>	<i>aac(6')/Ib-cr, aadA5</i>	FII, FIB	F2A1-:B-	BS512, IMG531	15	- +
			CTX-M-15	C2	<i>gyrA1AB</i>	<i>parC1aAB</i>	<i>aac(6')/Ib-cr, aadA5</i>	FII, FIB	F2A1-:B-	BS512	15	- +

^aInc, incompatibility; ND, not described so far; +, +, early biofilm producer; -, delayed biofilm producer; -, never a biofilm producer.

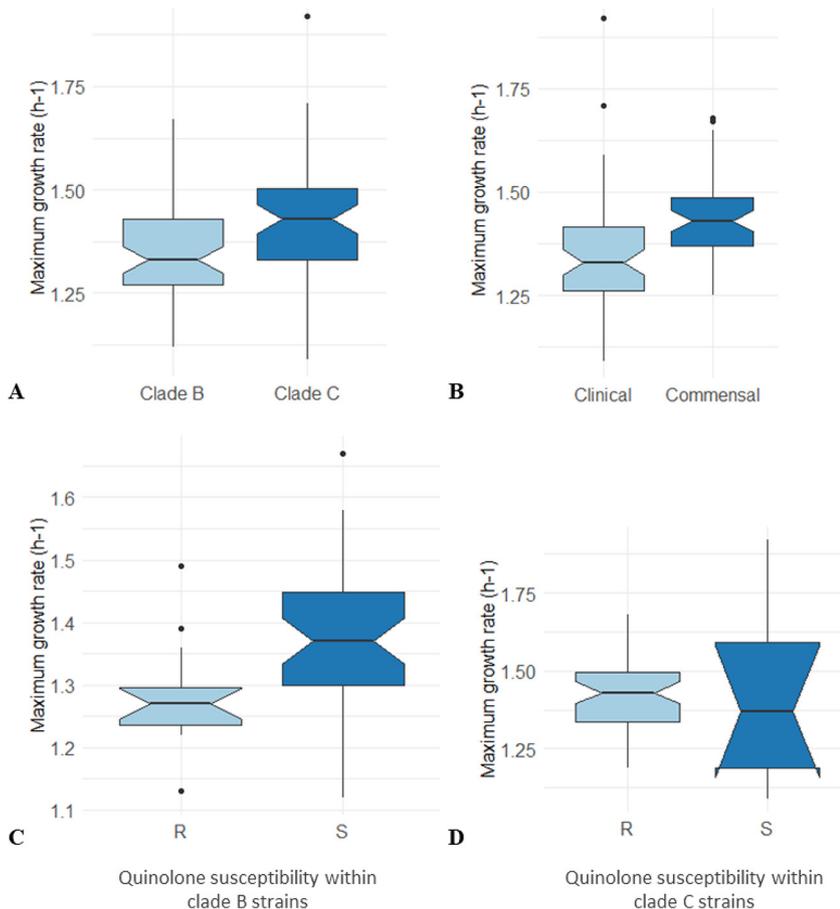


FIG 2 Maximum growth rate (MGR) in LB of the 39 O25b:H4 ST131 *Escherichia coli* strains studied. MGR was calculated from three independent LB culture assays and is expressed in hours⁻¹. Box plots indicate the MGR distribution across strains, and horizontal black bars indicate median values. The upper and lower ends of the box correspond to the upper and the lower quartiles, respectively. Error bars represent the standard error of the mean from three experiments. Outliers are represented by dots. (A) MGR of clade B and clade C strains. (B) MGR of clinical and commensal (fecal) strains. (C) MGR of nalidixic acid-resistant (R) and -susceptible (S) strains within clade B. (D) MGR of nalidixic acid-resistant (R) and -susceptible (S) strains within clade C. No significant difference was observed in each comparison (Wilcoxon test).

In sum, the MGR in a complex medium did not differ significantly between clade B and clade C strains, whereas clade B strains had a higher capability than clade C strains to form a biofilm early and express type 1 and curli fimbriae earlier.

Potential role of the *fimB* and the *ibeA* genes in the different phenotypes of early biofilm formation and type 1 fimbria expression identified among the 39 O25b:H4 ST131 strains studied. To decipher the molecular supports for the different early biofilm formation and type 1 fimbria expression phenotypes observed among the 39 strains, we analyzed the *fim* operon across the 39 studied strains, knowing that (i) the *ISEc55* insertion sequence, which encodes a cofactor of type 1 fimbria synthesis regulation, has been described within the clade C *fimB* gene (21) and (ii) Recombinant displayed the *fimH30* allele previously thought to be a specific trait of clade C strains. Between the *fim* operon of the *fimH22* clade B strains and that of Recombinant, there were from 5 to 50 SNPs per gene, depending on the gene (Fig. 4). Inversely, between the *fim* operon of Recombinant and that of clade C strains, there was a perfect match, except for two SNPs and the absence of *ISEc55* within the *fimB* gene of Recombinant (Fig. 4). We also found a partial deletion of the *fim* operon in two *fimH22* clade B strains, the B4 subclade strain 001-001 and the B5 subclade strain H1447 (data not shown).

To assess the role of the *fimB* gene in early biofilm formation and type 1 fimbria expression, we constructed Δ *fimB* mutants from the B1 subclade strain S250, which

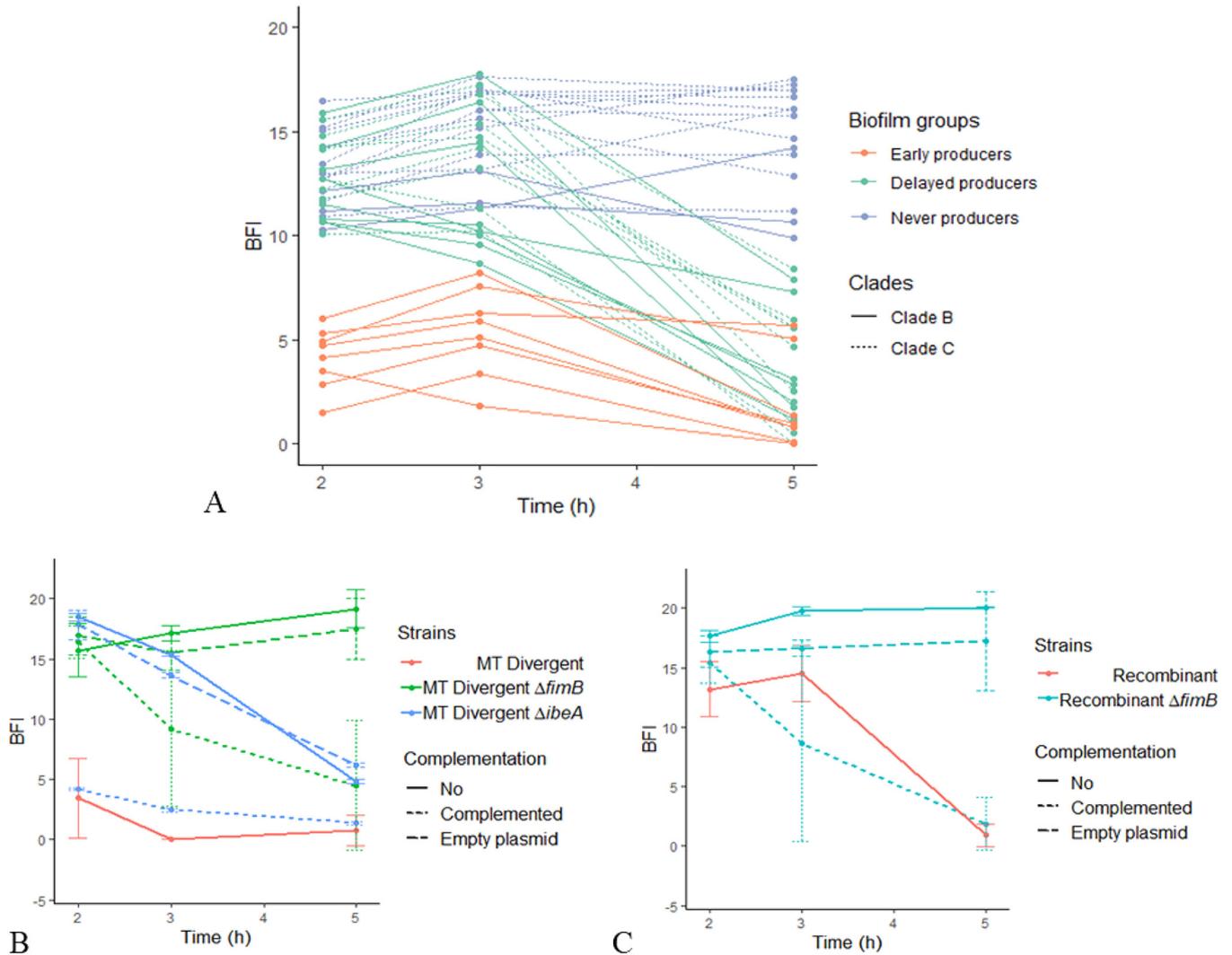


FIG 3 Early biofilm kinetics of the 39 O25b:H4 ST131 *Escherichia coli* strains and mutants studied. The BioFilm ring test was used to measure early biofilm formation, i.e., after 2, 3, and 5 h of static incubation. Biofilm index (BFI) values are inversely proportional to the biofilm formation capacity, with a BFI value of ≤ 10 meaning biofilm formation and a BFI value of > 10 meaning no biofilm formation. (A) Biofilm kinetics for the 39 O25b ST131 strains. Clade B strains were more frequently early biofilm producers than clade C strains ($P < 0.0001$). (B) Biofilm kinetics for (i) MT Divergent and its $\Delta fimB::FRT$ and $\Delta ibeA::FRT$ mutants, which became never and delayed biofilm producers, respectively, in the 5-h assay period; (ii) MT Divergent $\Delta fimB::FRT$ and $\Delta ibeA::FRT$ mutants complemented with the wild-type *fimB* gene ($\Delta fimBc$) and the wild-type *ibeA* gene (*ibeAc*), both of which were harbored by plasmid pSC-A, in which the early biofilm formation phenotype of the parental strain was partially and totally restored, respectively; and (iii) MT Divergent $\Delta fimB::FRT$ and MT Divergent $\Delta ibeA::FRT$ mutants complemented with the empty plasmid pSC-A ($\Delta fimBp$ and $\Delta ibeAp$, respectively). (C) Biofilm kinetics of Recombinant, its $\Delta fimB::FRT$ mutant, and the Recombinant $\Delta fimB::FRT$ mutant complemented with the wild-type *fimB* gene harbored by plasmid pSC-A ($\Delta fimBc$) or transfected with the empty plasmid pSC-A ($\Delta fimBp$). The $\Delta fimB::FRT$ mutant lost the capacity to produce a biofilm early, whereas when it was complemented with the wild-type *fimB* gene, it produced a biofilm earlier than the parental strain.

displayed an early (from 2 h) phenotype for the two traits (Tables 1 and 4), and Recombinant, which displayed a delayed (at 5 h) phenotype for the two traits (Tables 1 and 4). Both the S250 $\Delta fimB$ and Recombinant $\Delta fimB$ mutants lost their ability to form a biofilm (Fig. 3B and C) and to express type 1 fimbriae during the 5-h time period (Table 4). Complementation with the wild-type S250 *fimB* gene rendered the complemented strain a producer of a biofilm from 3 h, but at a level lower than that observed in strain S250 (Fig. 3B), and restored the expression of type 1 fimbriae observed in strain S250 (Table 4). Complementation with the wild-type recombinant *fimB* gene rendered the complemented strain a producer of a biofilm from 3 h, meaning that it was an earlier producer than Recombinant (Fig. 3C), and permitted the expression of type 1 fimbriae during the 5-h period of the experiment instead of only at 3 h, as observed in Recombinant (Table 4). Type 1 fimbria expression was also assessed after 24 h of

TABLE 2 Yeast agglutination test applied to the 39 O25b:H4 *Escherichia coli* strains under different growth conditions and at different time points according to biofilm production phenotype and clade type^a

Group	No. (%) of strains positive by yeast agglutination test			
	Early test, shaking culture		Standard test, 24 h ^a	
	2 h	5 h	Shaking culture	Static culture
Biofilm ++ (n = 8)	8 (100) ^{b,c}	8 (100) ^{b,c}	8 (100) ^{b,c}	8 (100)
Biofilm -+ (n = 16)	5 (31) ^{b,c}	5 (31) ^{b,c,d}	4 (25) ^{b,c}	11 (69)
Biofilm - (n = 15)	2 (13) ^c	0 ^{c,d}	0 ^c	11 (73)
Clade B (n = 19)	15 (79) ^c	13 (68) ^c	12 (63) ^c	17 (90)
Clade C (n = 20)	0 ^c	0 ^c	0 ^c	13 (65)

^aThe standard test was performed as described by Totsika et al. (21). Early tests were always negative in static cultures (data not shown). ++, early biofilm producer; -+, delayed biofilm producer; -, never biofilm producer. *P* values were determined by Fisher's exact test.

^b*P* < 0.01.

^c*P* < 0.0001.

^d*P* < 0.05.

incubation (Table 4). This experiment showed that the S250 Δ *fimB* mutant did not express type 1 fimbriae at 24 h under either shaking or static conditions, whereas the Recombinant Δ *fimB* mutant and the tested clade C strain (CES164C) did, but only under static conditions (Table 4).

We then investigated the *ibeA* gene, knowing that (i) it was previously suggested that the brain endothelial cell invasion and adhesion determinants encoded by the *ibeA* gene and the *ibeT* gene, respectively, would have a role in modulating the expression of type 1 fimbriae (22) and (ii) the *ibeA* gene is absent in clade C strains (Table 5). We found that the *gimA* locus, including the *ibeAT* genes, encompassed with the *ibeR* gene in the *gimA4* operon, was present in all clade B strains except for strain 001-001 (data not shown) and Recombinant, in which it was completely deleted, as in clade C strains (Fig. 4). Additionally, we found that the inactivation of the *ibeA* gene of strain S250 rendered the S250 Δ *ibeA* mutant a delayed producer of a biofilm, whereas its complementation with the wild-type *ibeA* gene restored the early biofilm phenotype observed in the parental strain (Fig. 3B). Inversely, *ibeA* inactivation in strain S250, as well as that of the *ibeART* operon, had no impact on the expression of type 1 fimbriae (Table 4).

In sum, the targeted-gene analysis performed *in vitro*, i.e., in a simple environment, showed the involvement of both H22/H30 FimB and IbeA in early biofilm formation and only that of H22/H30 FimB in the early expression of type 1 fimbriae.

***In vivo* virulence and ExPEC VF gene content of the 39 studied O25b:H4 ST131 strains.** By using a well-calibrated mouse sepsis model (23), we assessed, *in vivo*, a more complex phenotype i.e., the intrinsic extraintestinal virulence, of the 39 strains studied. We found that clade B strains were significantly more virulent than clade C strains (*P* = 2e-10) (Fig. 5). Among the clade B strains (Fig. 6A), no significant difference was

TABLE 3 Colony color and morphotype on Congo red agar plates of the 39 O25b:H4 ST131 *Escherichia coli* strains according to biofilm production phenotype and clade type^a

Group	No. (%) of strains with a positive Congo red test	
	Congo red staining	rdar morphotype
Biofilm ++ (n = 8)	8 (100) ^b	8 (100) ^b
Biofilm -+ (n = 16)	1 (6) ^b	1 (6) ^b
Biofilm - (n = 15)	4 (27) ^b	3 (20) ^b
Clade B (n = 19)	9 (47) ^c	9 (47) ^c
Clade C (n = 20)	3 (15) ^c	2 (10) ^c

^a++, early biofilm producer; -+, delayed biofilm producer; -, never biofilm producer; rdar, red, dry, and rough colonies. *P* values were determined by Fisher's exact test.

^b*P* < 0.0001 for the early biofilm producer group versus both the delayed biofilm producer and never biofilm producer groups.

^c*P* < 0.001 for clade B versus clade C.

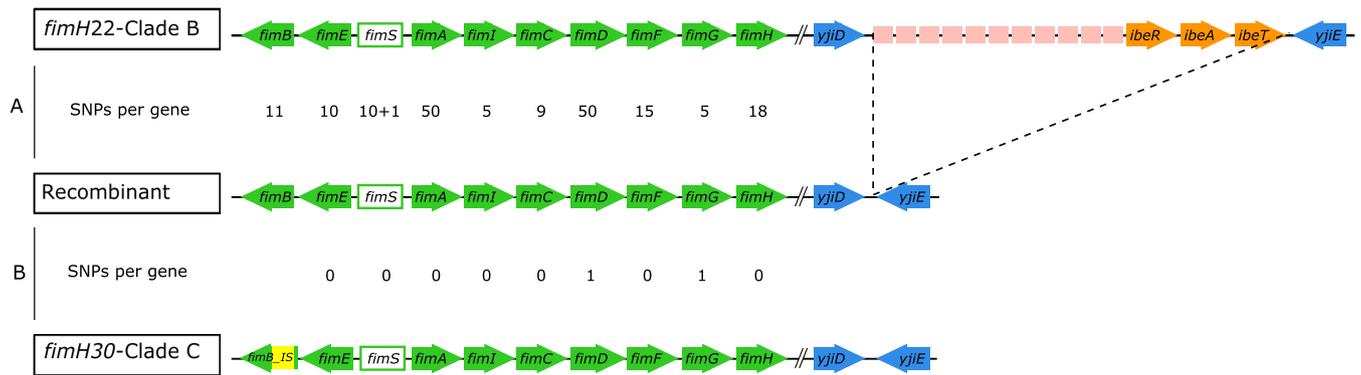


FIG 4 Schematic representation of the *fim* operon and the *gimA* locus in O25b:H4 ST131 *fimH22*-clade B strains, Recombinant, and clade C strains. The genes of the *fim* operon are depicted by green arrows, except for the *fimS* gene, which is an invertible element; those of the *gimA4* operon, including the *ibeART* operon, which are indicated by other arrows; and those of the three other operons of the *gimA* locus, which are indicated by pink rectangles. The dashed lines indicate the deletion of the *gimA* locus in both Recombinant and the clade C strains. Yellow indicates the ISEc55 insertion within the *fimB* gene of all clade C strains. (A) Comparison between *fimH22*-clade B strains based on a B1 subclade strain (strain S250) and Recombinant (B4 subclade). (B) Comparison between Recombinant and clade C strains based on the operon of a C1 subclade strain (CES164C). SNPs, single nucleotide polymorphisms; +1, the presence of a gap.

observed between the strains of each subclade except for B3 subclade strains, which killed mice significantly faster than those of B4 subclade ($P = 0.02$). Among the clade C strains (Fig. 6B), the C0 subclade strains were significantly more virulent than those of the C1 subclade ($P = 5e-09$) and C2 subclade ($P = 6e-05$), while there was no significant difference between those of the C1 and C2 subclades.

Considering the described association between virulence assessed in mouse sepsis models and the number of genes encoding ExPEC virulence factors (VFs) (24, 25), we analyzed the distribution of the ExPEC VF-encoding genes in the 39 strains. The average number of VF-encoding genes was higher in clade B strains (16.2 ± 2.4) than in clade C strains (13.3 ± 1.3) ($P = 4e-07$) (Table 5). The *hlyF*, *cdtB*, *iroN*, *kpsMIII*, *cvaC*, *iss*, and *ibeA* genes were significantly more frequent among clade B strains than among clade C strains, whereas the opposite was found for the *sat* gene (Table 5). Within clade B, there was no significant difference between the mean number of VF genes of the

TABLE 4 Yeast agglutination test applied to B1 subclade strain S250 (MT Divergent) and its mutants, B4 subclade strain CES131 (Recombinant) and its mutant, and C1 subclade strain CES164 (Emergent)

Strain	Yeast agglutination test result ^a			
	Shaking growth			
	Early test ^b			Static growth, standard test at 24 h
At 2 h	At 5 h	Standard test at 24 h		
S250	++++	++++	+++	++++
S250 Δ <i>fimB</i> :: <i>kan</i>	-	-	-	-
S250 Δ <i>fimB</i> ::FRT/pSC-AS250 <i>fimB</i>	++++	++++	+++	+++
S250 Δ <i>fimB</i> :: FRT/pSC-A	-	-	-	-
S250 Δ <i>ibeA</i> :: <i>kan</i>	++++	++++	+++	++++
S250 Δ <i>fimB</i> ::FRT Δ <i>ibeA</i> :: <i>kan</i>	-	-	-	-
S250 Δ <i>ibeART</i> :: <i>kan</i>	++++	++++	+++	++++
S250 Δ <i>fimB</i> ::FRT Δ <i>ibeART</i> :: <i>kan</i>	-	-	-	-
CES131C	-	+	-	++++
CES131C Δ <i>fimB</i> :: <i>kan</i>	-	-	-	+++
CES131C Δ <i>fimB</i> ::FRT/pSC-ACES131C <i>fimB</i>	++	+++	+++	++++
CES131C Δ <i>fimB</i> ::FRT/pSC-A	-	+	-	+++
CES164C	-	-	-	+++
Positive control (<i>E. coli</i> UT189)	+++	+++	+++	+++
Negative control (<i>E. coli</i> UBA83972)	-	-	-	-

^aThe yeast agglutination test was positive when agglutination appeared within 3 min. +++++, fast appearance of agglutination with thick aggregations; +++, fast appearance of agglutination with fine aggregations; ++, slow appearance of agglutination with thick aggregations; +, slow appearance of agglutination with fine aggregations; -, no agglutination.

^bThe early test was always negative under static growth conditions (data not shown).

TABLE 5 Main virulence factor-encoding genes in the 39 O25b:H4 ST131 *Escherichia coli* strains, by clade

Gene	No. (%) of strains			P value ^a for clade B vs clade C
	Total (n = 39)	Clade B (n = 19)	Clade C (n = 20)	
Adhesin				
<i>yfcV</i>	39 (100)	19	20	
<i>fimH</i>	39 (100)	19	20	
<i>papAH</i>	2 (5)	2	0	
<i>papC</i>	2 (5)	2	0	
<i>papEF</i>	3 (8)	2	1	
<i>sfa</i> and <i>foc</i>	1 (3)	0	1	
<i>afa</i> and <i>dra</i>	8 (21)	6	2	
Toxin				
<i>hlyA</i>	0	0	0	
<i>hlyF</i>	10 (26)	10	0	0.0002
<i>cnf1</i>	0	0	0	
<i>cdtB</i>	7 (18)	7	0	0.003
<i>sat</i>	25 (64)	7	18	0.0008
<i>vat</i>	0	0	0	
<i>tsh</i>	1 (3)	1	0	
Iron uptake				
<i>iroN</i>	10 (26)	10	0	0.0002
<i>fyuA</i>	39 (100)	19	20	
<i>iutA</i>	32 (82)	14	19	
<i>iucD</i>	32 (82)	14	19	
<i>irp2</i>	39 (100)	19	20	
<i>chuA</i>	39 (100)	19	20	
Capsule				
<i>kpsMIII</i>	33 (85)	19	14	0.02
<i>kpsMIII</i>	0	0	0	
<i>K1</i>	3 (8)	3	0	
<i>K2</i>	0	0	0	
<i>K5</i>	30 (77)	16	14	
Miscellaneous				
<i>iss</i> type 1	11 (28)	11	0	0.00005
<i>traT</i>	36 (92)	17	19	
<i>cvaC</i>	5 (13)	5	0	0.02
<i>ibeA</i>	17 (44)	17	0	3×10^{-9}
<i>usp</i>	39 (100)	19	20	
<i>ompT</i>	39 (100)	19	20	
<i>malX</i>	39 (100)	19	20	
Mean \pm SD VF ^b	14.9 \pm 2.5	16.6 \pm 2.4	13.3 \pm 1.3	0.000004

^aP values (shown when they were <0.05) were obtained by Fisher's exact test for the gene distribution comparison between clades and by the Wilcoxon rank-sum test for the VF mean comparison.

^bVF, virulence factor-encoding gene.

strains of the different B subclades, except for those belonging to the B3 subclade, which displayed a higher mean number of VF genes than those belonging to the B4 subclade ($P < 0.05$) (data not shown). Within clade C, there was no significant difference either between the C1 and C2 subclade strains or between those of the C0 and C2 subclades, but there was a significant difference between those of C0 and C1 subclades, with the C0 subclade strains having a higher mean number of VF genes ($P = 0.03$) (data not shown). Among the four main VF-based virotypes (virotypes A, B, C and D) previously described for O25b:H4 ST131 (26), we found that virotypes D and C were significantly associated with clade B and clade C strains, respectively ($P < 0.00001$), and that two B4 subclade strains (including Recombinant), one C1 subclade strain, and three C2 subclade strains displayed previously undescribed virotypes (Table 1).

In sum, clade B strains displayed a higher *in vivo* virulence than clade C strains in a mouse sepsis model in relation to a greater number of VFs.

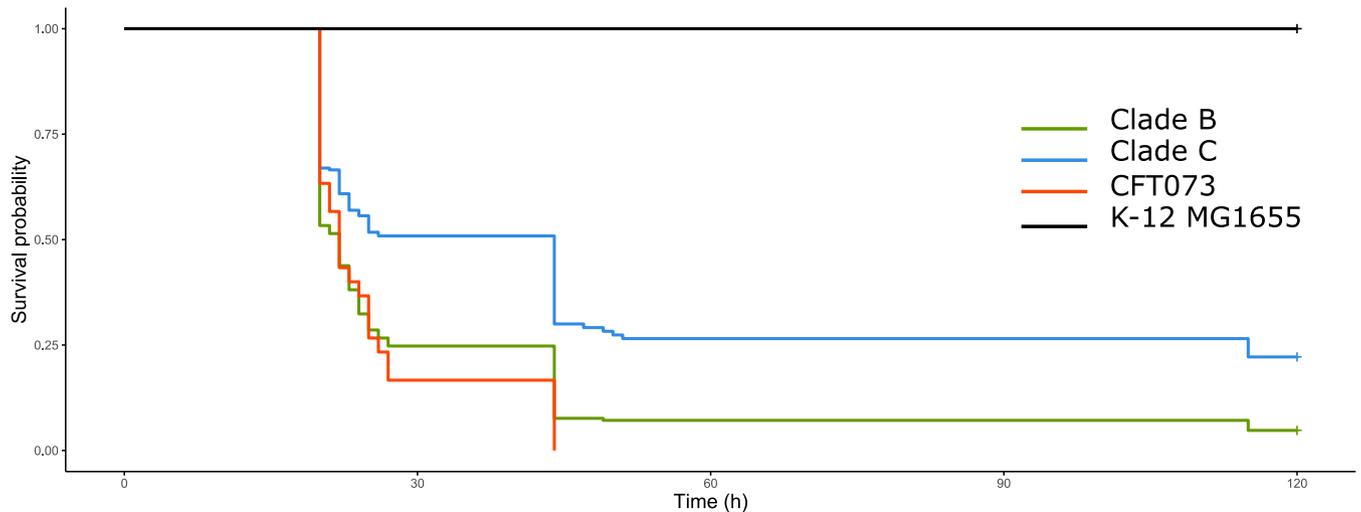


FIG 5 Kaplan-Meier survival curves of mice with sepsis due to the 39 O25b:H4 ST131 *Escherichia coli* strains studied, by clade. Clade B and C strains and controls (positive control, strain CFT073; negative control, strain K-12 MG1655) are represented by different colored lines. Clade B strains (210 mice) killed mice significantly faster than clade C strains (230 mice) ($P = 2e-10$).

Selection of strains for further *in vivo* experiments. To go deeper into the differences observed in the *in vitro* and *in vivo* virulence phenotypes between clades B and C, we selected three strains that allowed us to test (i) gene inactivation and (ii) other relevant *in vivo* mouse model complex phenotypes. Thus, within clade B, we selected strain S250, belonging to the B1 subclade, whose fitness had already been assessed in *Caenorhabditis elegans* and zebrafish models (27) and *in vitro* (28) and which corresponded, in our collection, to one of the two most anciently diverged strains. We called this strain “MT Divergent,” where MT stands for “the most.” Among the C1 and C2 subclade strains that did not show significant differences with regard to early biofilm formation, the mean number of VF genes, and mouse lethality, we selected C1 subclade strain CES164C, which we called “Emergent.” We also retained Recombinant. Table 6 presents the major characteristics displayed by these three strains. To assess *in vivo* the potential role of the *fimB* gene and the *ibeART* operon, we retained the $\Delta fimB::kan$, $\Delta ibeART::kan$, and $\Delta fimB::FRT \Delta ibeART::kan$ recombination target (FRT) $\Delta ibeART::kan$ mutants of MT Divergent and the $\Delta fimB::kan$ mutant of Recombinant.

Competition assays in mouse models of sepsis, intestinal colonization, and urinary tract infection (UTI). Before performing competition assays between MT Divergent, Emergent, and Recombinant and between parental strains and mutants in the mouse sepsis model, we checked their virulence phenotype in this model in a mono-infection (Fig. S3). MT Divergent killed mice faster than Emergent ($P = 0.04$) but not than Recombinant ($P = 0.8$), whereas Recombinant and Emergent did not kill mice differently ($P = 0.06$) (Fig. S3A). There was no significant difference either between MT Divergent and its $\Delta fimB::kan$, $\Delta ibeART::kan$, and $\Delta fimB::FRT \Delta ibeART::kan$ mutants or between Recombinant and its $\Delta fimB::kan$ mutant (Fig. S3B).

In competition assays performed in the mouse sepsis model, MT Divergent outcompeted both Recombinant ($P = 0.002$) and Emergent ($P = 0.002$) and Recombinant outcompeted Emergent ($P = 0.002$) (Fig. 7). MT Divergent did not outcompete its $\Delta fimB::kan$ mutant, whereas Recombinant outcompeted its $\Delta fimB::kan$ mutant ($P = 0.004$) (Fig. 7). Inversely, MT Divergent outcompeted its $\Delta ibeART::kan$ ($P = 0.01$) and $\Delta fimB::FRT \Delta ibeART::kan$ ($P = 0.002$) mutants (Fig. 7). Neither the kanamycin resistance cassette $\Delta fimB$ ($P = 0.3$) nor the kanamycin resistance cassette $\Delta ibeART$ ($P = 0.06$) had a cost to *in vivo* virulence in this mouse model (Fig. 7).

Competitions between the three strains as well as between the parental strains and their mutants were also performed in a mouse intestinal colonization model because intestinal colonization is the first step of infection and extraintestinal virulence is a

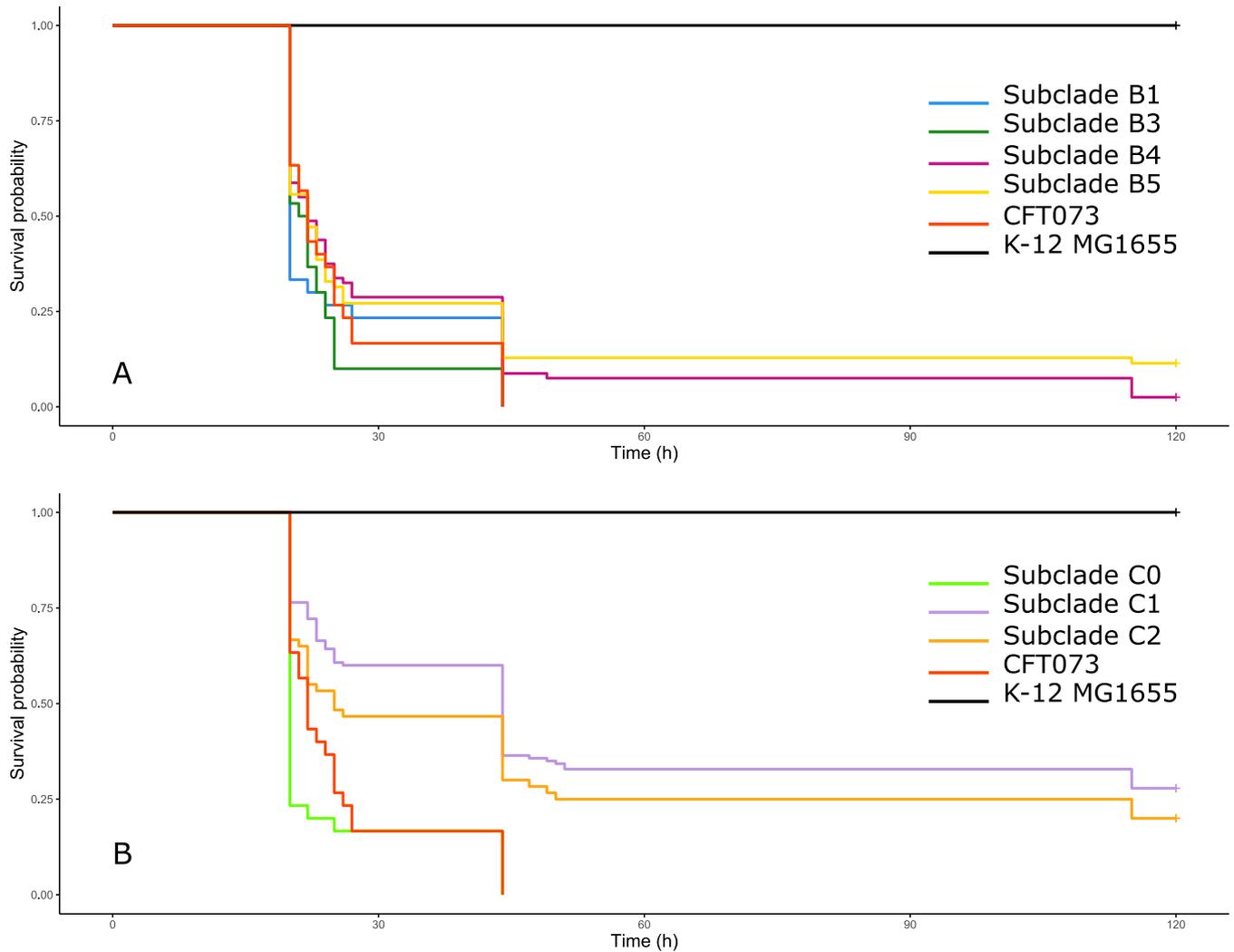


FIG 6 Kaplan-Meier survival curves of mice with sepsis due to the 39 O25b:H4 ST131 *Escherichia coli* strains studied, by subclade. B subclade strains (A), C subclade strains (B), and controls (positive control, strain CFT073; negative control, strain K-12 MG1655) are represented by different colored lines. No significant difference was observed between B1 subclade strains (30 mice) and those of the B3 subclade (30 mice), B4 subclade (80 mice), and B5 subclade (70 mice), whereas B3 subclade strains killed mice significantly faster than those of the B4 subclade ($P = 0.02$). Among the clade C strains, C0 subclade strains (30 mice) killed mice significantly faster than those of the C1 subclade (140 mice) ($P = 5e-09$) and C2 subclade (60 mice) ($P = 6e-05$), while there was no significant difference between the C1 and C2 subclade strains.

coincidental by-product of commensalism in B2 phylogenetic group *E. coli* strains (29). Given that this model requires competitive strains with an equal level of susceptibility to streptomycin and that both MT Divergent and Emergent were susceptible to streptomycin, we used a kanamycin resistance cassette to inactivate the plasmid-mediated *aadA2* gene, encoding streptomycin resistance, in Recombinant. This cassette showed no fitness cost through a competition assay between Recombinant and Recombinant $\Delta aadA2::kan$ under planktonic conditions in LB (data not shown). In this mouse model, we found that MT Divergent outcompeted Emergent ($P = 0.02$) at day 4 and day 7, whereas there was no significant difference at any time point between MT Divergent and Recombinant or between Recombinant and Emergent (Fig. 8). There was also no significant difference between the mutants and their parental strains, i.e., the $\Delta fimB::kan$, $\Delta ibeART::kan$, and $\Delta fimB::FRT \Delta ibeART::kan$ mutants for MT Divergent and the $\Delta fimB::kan$ mutant for Recombinant (Fig. 8).

Finally, we put MT Divergent, Recombinant, and Emergent in competition in the mouse UTI model, knowing that clade C strains have frequently been identified among uropathogenic *E. coli* (UPEC) strains (7). We found that MT Divergent outcompeted

TABLE 6 Comparative characteristics of the three O25b:H4 *Escherichia coli* strains selected for competition assays^a

Strain	MGR (h ⁻¹)	Colicin/phage production	Biofilm formation phenotype	Time (h) of type 1 fimbria expression	Congo red assay phenotype	Antibiotic resistance-encoding genes	No. of VF-encoding genes ^b	No. of ExPEC VF-encoding genes	<i>fim</i> operon type	<i>gimA</i> locus
MT Divergent (B1 subclade S250)	1.35 ± 0.06	No	Early	2 ^c	rdar		53	14	H22	+
Emergent (C1 subclade CES164C)	1.48 ± 0.06	No	Never	24 ^d	Red staining	<i>blq_{TEM}</i> , <i>gyrA1AB</i> , <i>parC1aAB</i>	52	13	H30	-
Recombinant (B4 subclade CES131C)	1.47 ± 0.09	No	Delayed	5 ^d	White staining	<i>blq_{TEM}</i> , <i>aac(3)-IId</i> , <i>aadA2</i> , <i>mph(A)</i> , <i>sul1</i> , <i>tet(A)</i> , <i>dfrA12</i>	63	15	H30	-

^aVF, virulence factor; rdar, red, dry, and rough; +, present.

^bSee Fig. S2 in the supplemental material for details.

^cWith shaking growth.

^dWith static growth.

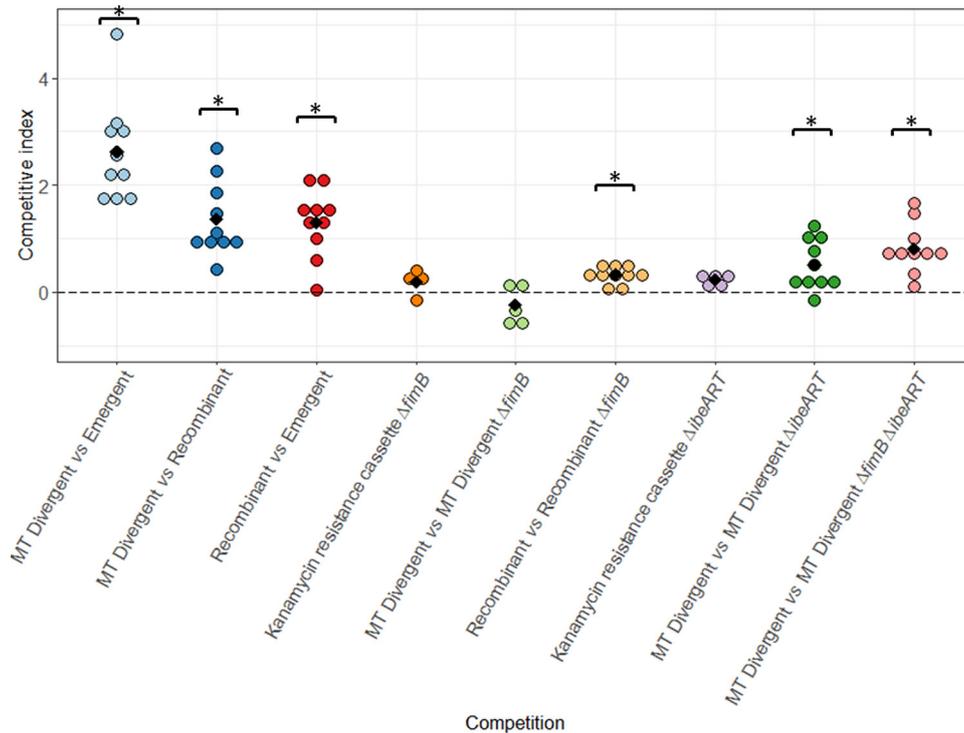


FIG 7 Competitions between MT Divergent, Recombinant, and Emergent and between MT Divergent and its mutants and Recombinant and its mutant in the mouse sepsis model. Competitions were determined by the spleen bacterial load. The competitive index (CI) is expressed in log. A CI above 0 means that the isolate cited first in the legend outcompetes the one cited second, and a CI below 0 means that the isolate cited second outcompetes the one cited first. Each dot represents one mouse, and different colors were attributed to the different competitions. Black diamonds depict the mean values of CI. MT Divergent outcompeted Emergent ($P = 0.002$) and Recombinant ($P = 0.002$), and Recombinant outcompeted Emergent ($P = 0.002$). Before performing competitions between the parental strains and their mutant(s), we checked that the kanamycin resistance cassette introduced in the targeted gene did not have any cost: kanamycin resistance cassette Δ fimB (competition between MT Divergent Δ fimB::FRT versus MT Divergent Δ fimB::kan and Recombinant Δ fimB::FRT versus Recombinant Δ fimB::kan, $P = 0.2$) and kanamycin resistance cassette Δ ibeART (competition between MT Divergent Δ ibeART::FRT versus MT Divergent Δ ibeART::kan and MT Divergent Δ fimB::FRT Δ ibeART::FRT versus MT Divergent Δ fimB::FRT Δ ibeART::kan, $P = 0.062$). *, significant difference generated by the Wilcoxon signed-rank test. MT Divergent did not outcompete MT Divergent Δ fimB::kan ($P = 0.3$), whereas Recombinant outcompeted Recombinant Δ fimB::kan ($P = 0.004$). MT Divergent outcompeted both MT Divergent Δ ibeART::kan ($P = 0.001$) and MT Divergent Δ fimB::FRT Δ ibeART::kan ($P = 0.002$).

Emergent in the bladder ($P = 0.004$) and the kidneys ($P = 0.004$) and that Recombinant outcompeted Emergent in the bladder ($P = 0.008$) but not in the kidneys ($P = 0.13$), whereas there was no significant difference between MT Divergent and Recombinant either in the bladder ($P = 0.08$) or in the kidneys ($P = 0.05$) (Fig. 9). In sum, MT Divergent outcompeted Emergent regardless of the mouse model used.

DISCUSSION

To find out the factors other than multidrug resistance that could have favored the dissemination of *E. coli* O25b:H4 ST131 clade C, we used an original approach based on phenotypic comparisons of clade B and clade C strains. To this end, we established an O25b:H4 ST131 collection composed of clade B ($n = 19$) and clade C ($n = 20$) strains belonging to the various B and C subclades. However, in this collection, we found a B4 subclade strain with a *fimH30* allele (called "Recombinant"), a peculiar feature previously found in another B4 subclade strain (5). Here, we clarified that Recombinant shared with clade C not only the same *fim* operon (however, it lacked the *ISEc55* sequence always found in the clade C *fimB* gene) but also a deletion of the *gimA* locus, including the *ibeART* operon. This suggests that identical genetic events had occurred in clades B and C but, apparently, with a lower adaptive success in the genetic background of clade B than in that of clade C.

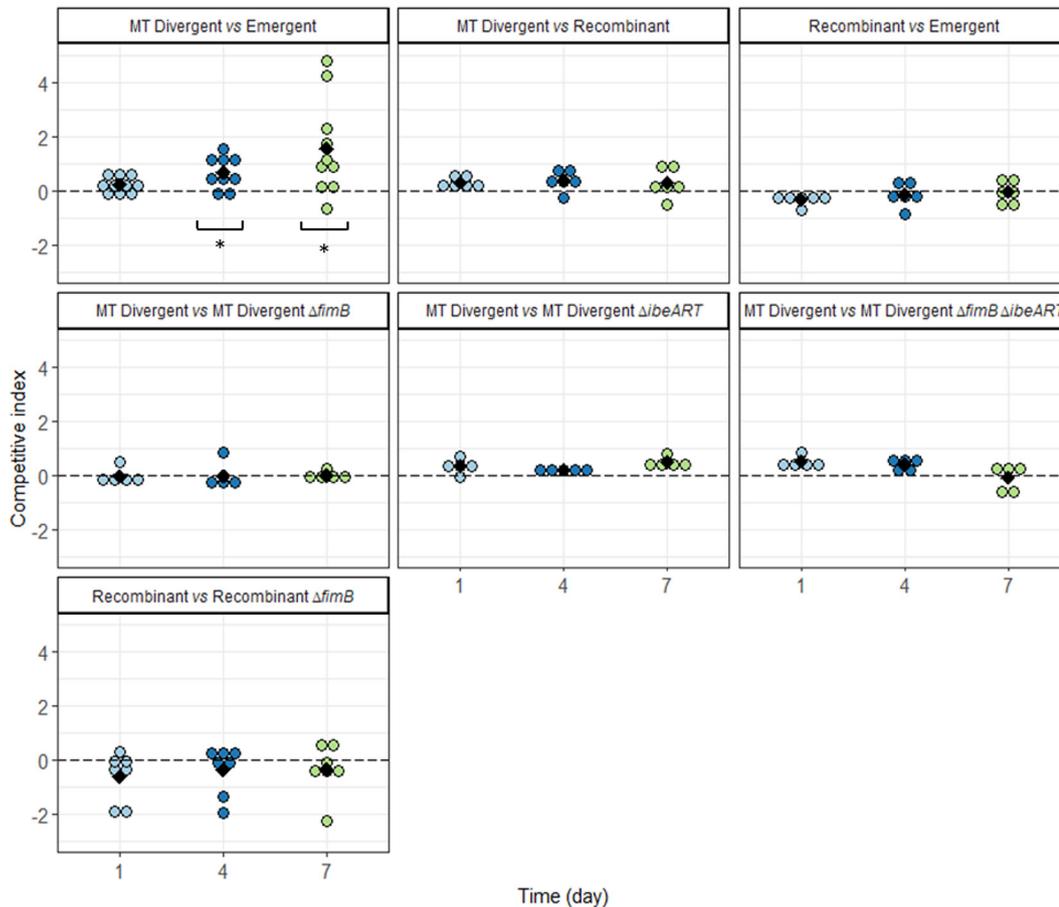


FIG 8 Competitions between MT Divergent, Recombinant, and Emergent and between MT Divergent and its mutants and Recombinant and its mutant in the mouse gut colonization model. Competitions were determined by the fecal bacterial load. The competitive index (CI) is expressed in log. A CI above 0 means that the isolate cited first in the legend outcompetes the one cited second, and a CI below 0 means that the isolate cited second outcompetes the one cited first. Each dot represents one mouse, with the different colors indicating the time period, as follows: light blue, day 1 postinoculation; dark blue, day 4 postinoculation; green, day 7 postinoculation. Black diamonds depict the mean values of CI. All mutants harbored the *kan* gene of the kanamycin resistance cassette. *, significant difference generated by the Wilcoxon signed-rank test and corrected for multiple comparisons by the Benjamini-Hochberg procedure. MT Divergent outcompeted Emergent at both day 4 ($P = 0.01$) and day 7 ($P = 0.001$), whereas the other competitions were not significant (P values, between 1 and 0.09, according to the two strain types in the competition and the competition day).

In all but one (MGR in LB) phenotypic experiment applied to the 39 strains, we found that clade B had a significantly higher virulence and competitive ability than clade C, and the level of these features for Recombinant was intermediate between those of *fimH22*-clade B and clade C strains. This indicated that, despite the same ability to grow under optimal conditions *in vitro*, the strains behaved differently in more complex environments more closely related to the normal niches of *E. coli* ST131.

We confirmed that the early biofilm production phenotype, which is a property displayed by <10% of ExPEC strains belonging to few *E. coli* lineages (19), was always displayed by clade B strains and never by clade C strains (19, 28). By using the MT Divergent $\Delta H22$ -*fimB* and Recombinant $\Delta H30$ -*fimB* mutants, we showed that biofilm formation and type 1 fimbriae expression during the first 5 h of incubation require a functional *fimB* gene. We also showed that the Recombinant $H30$ - Δ *fimB* mutant required, like Emergent, at least 24 h of growth under static conditions to express type 1 fimbriae, whereas the MT Divergent $\Delta H22$ -*fimB* mutant was unable to express these fimbriae at 24 h under both shaking and static conditions. All these findings strongly suggest that the presence of the ISEc55 sequence in the *fimB* gene of clade C strains contributes to their incapacity to form a biofilm early and express type 1 fimbriae

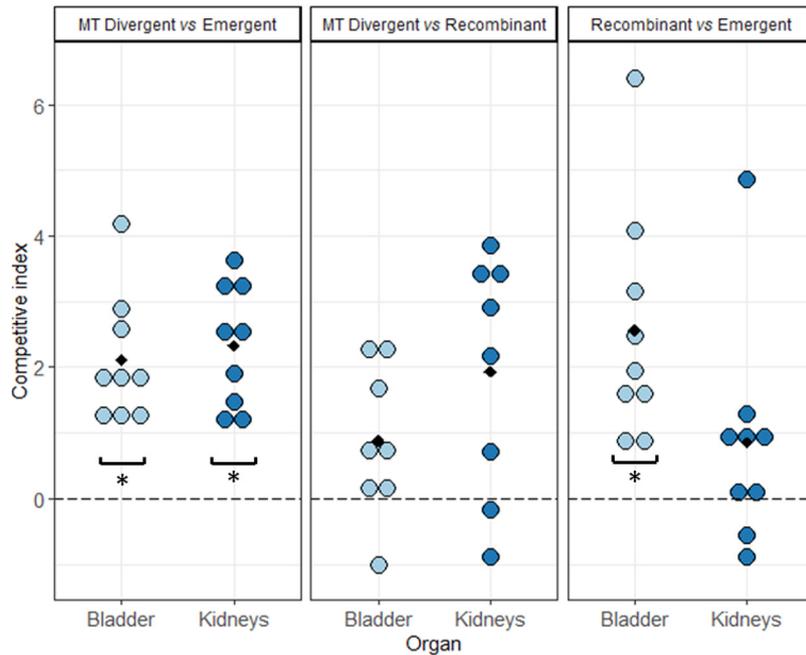


FIG 9 Competitions between MT Divergent, Recombinant, and Emergent in the mouse urinary tract infection model. Competitions were determined by the bladder and kidney bacterial loads. Each dot represents one mouse, with the different colors indicating the organs tested; light blue for urine and dark blues for the kidneys. The competitive index (CI) is expressed in log. A CI above 0 means that the isolate cited first in the legend outcompetes the one cited second, and a CI below 0 means that the isolate cited second outcompetes the one cited first. Black diamonds depict the mean values of CI. *, significant difference generated by the Wilcoxon signed-rank test. MT Divergent outcompeted Emergent in the bladder ($P = 0.004$) and the kidneys ($P = 0.004$), whereas it did not outcompete Recombinant either in the bladder ($P = 0.08$) or in the kidneys ($P = 0.05$). Recombinant outcompeted Emergent in the bladder ($P = 0.004$) but not in the kidneys ($P = 0.13$).

before 24 h of incubation and that, in addition to the genetic differences between the two *fim* operon types, there could be differences in their regulation. Our results for type 1 fimbria expression by Emergent and other clade C strains are concordant with those found by Totsika et al. with clade C strain EC958 and other O25b:H4 ST131 strains harboring the *ISEc55* sequence in the *fimB* gene (21). Using a mouse UTI model, these authors showed that strain EC958 colonized the bladder when type 1 fimbria expression was strong and persisted in urine when it was weak (21). On the other hand, Gunther et al., also through competition assays in a mouse UTI model between a wild-type CFT073 strain and mutants expressing their type 1 fimbriae only weakly or strongly, concluded that type 1 fimbria phase variation contributes significantly to virulence early (24 h postinoculation) in the 76-h course of the mouse UTI and profoundly influences colonization of the bladder (30). Our competition assays, performed in the same mouse UTI model as that used in the cited studies, showed that MT Divergent outcompeted Emergent in the bladder and the kidneys. These *in vivo* results, combined with those that we obtained *in vitro*, strongly suggest that the early expression of type 1 fimbriae allows clade B to quickly cause bladder colonization and pyelonephritis, while clade C requires more time to have a functional temporal on/off regulation of type 1 fimbria expression in a link to bladder colonization and urine persistence, respectively, as described for multiple UPEC strains by Schwan and Ding (31). In short, the loss of early expression of type 1 fimbriae by clade C strains makes them classical phylogroup B2 UPEC strains with, however, a peculiar trait, i.e., antibiotic and even multiantibiotic resistance (32).

The other major significant phenotypic differences between clade B and clade C consisted of a higher VF score and faster killing of mice by clade B than by clade C. This result is in accordance with the results of previous studies that showed the association

between virulence, as assessed in mouse sepsis models, and the number of VFs (24, 25). Within clade B, B3 subclade strains had both higher VF scores and a higher *in vivo* virulence than B4 subclade strains, whereas within clade C, C0 subclade strains had higher VF scores and a higher *in vivo* virulence than either C1 or C2 subclade strains, but this difference was not found between C1 and C2 subclade strains. Making comparisons between our results and those previously obtained with the same mouse sepsis model is difficult because the previous studies used only surrogates of clade B and clade C strains (strains with susceptibility to fluoroquinolones, *fimH30* versus non-*fimH30* alleles, and virotypes) (33–35). Using well-characterized strains, as we did, permitted the various levels of *in vivo* virulence between and within the clades and subclades of O25b:H4 ST131 to be deciphered.

We performed a genome-wide association study (GWAS) using Scoary software (36), which assesses the association between the presence/absence of genes and *in vivo* virulence in the mouse sepsis assay, but did not obtain any significant hit over the 17,951 tested genes (data not shown). This failure could have been due to the small number of strains tested; the limited genetic variability of the strains, all of which belonged to a single *E. coli* lineage (37); or an *in vivo* virulence resulting from VF cooperation in an additive manner to achieve extraintestinal virulence (25). However, targeted-gene analyses performed in competition assays in the mouse sepsis model showed that the *H30-fimB* gene was associated with the *in vivo* virulence of Recombinant and the *ibeART* operon was associated with that of MT Divergent. These results, together with those showing that MT Divergent outcompeted both Recombinant and Emergent and that Recombinant outcompeted Emergent in the mouse sepsis model, strongly suggest that the absence of the *ibeART* operon in Recombinant and clade C and the nonfunctionality of the *H30-fimB* gene in clade C may have contributed to the gradually lower levels of *in vivo* virulence displayed by Recombinant and Emergent in the mouse sepsis model.

As competition assays provide more evidence than assays with single strains of the difference in fitness of only a few percentage points, indicating stronger differences, we used only competition assays to study more thoroughly three representative strains (i.e., MT Divergent, Recombinant, and Emergent) in a mouse model of sepsis, gut colonization, and UTI. In all competition assays, Emergent was outcompeted by MT Divergent regardless of the mouse ecological niches, whereas it was outcompeted by Recombinant, depending on the niches: it was outcompeted in the blood and partially outcompeted in the urinary tract (in the bladder but not in the kidneys) but was not outcompeted in the gut. Such features may suggest an adaptation to ecological niches within the ST131 lineage, as assessed by McNally et al. (14).

In conclusion, by measuring several simple and more complex phenotypes, we highlighted the lower level of virulence of clade C strains than of clade B strains, what had previously been suggested in the zebrafish model (27).

It may appear to be surprising that the clade most disseminated worldwide, i.e., ST131 clade C, displays attenuated virulence. We propose two explanations for its widespread dissemination. First, the resistance to fluoroquinolones displayed by clade C may come at a direct cost (38). This cost would not be revealed in the MGR under planktonic conditions but would manifest *in vivo*, as suggested here in various mouse models. In spite of this cost, antibiotic resistance allowed clade C to spread. Second, the lower virulence may actually confer an additional evolutionary advantage to clade C. According to the trade-off theory (39), host exploitation by a pathogen evolves to an optimal level under a balance between the benefits in terms of transmission and the costs in terms of host mortality (40). Thus, lower virulence could confer improved fitness overall by avoiding severe infections (41, 42), even if it comes at the cost of a lower colonization ability.

Whether the phenotypic clade B and clade C differences represent a cost, in spite of which clade C is successful, or an additional adaptation remains an open question that requires further work, especially as the *in vivo* behavior of fluoroquinolone-susceptible Recombinant favors the additional adaptation hypothesis.

MATERIALS AND METHODS

More technical details for each section of Materials and Methods are available in Text S1 in the supplemental material.

Bacterial strains. The 39 O25b:H4 ST131 *E. coli* strains studied comprised 18 *fimH22* and 21 *fimH30* strains obtained between 1993 and 2012 from different geographic origins and sources (Table S1). The *E. coli* CFT073 and *E. coli* K-12 MG1655 strains were used as positive and negative controls, respectively, in the mouse sepsis model, and the *E. coli* UT189 and *E. coli* UBA83972 strains were used as positive and negative controls, respectively, in the yeast agglutination assays.

Genome sequencing and analysis. Whole-genome sequencing (WGS) of the 39 studied strains was performed (Table S2). All genomes were analyzed for plasmid content and typing (with the Plasmid Finder tool, which showed an identity of >95%, and plasmid multilocus sequence typing [pMLST] [43]). The Abricate tool (44) was used to detect (i) genes encoding antibiotic resistance with the ResFinder program (45) and (ii) genes encoding VF with a custom virulence database composed of VirulenceFinder (46), the virulence factor database (VFDB) (47), and the classical ExPEC VF-encoding genes (19). Virotypes were determined as previously described (26). All contigs were submitted to the MicroScope platform (48) for further gene investigations, such as the investigation of the *gyrA* and *parC* genes. When necessary, the presence or absence of some genes was controlled by PCR.

To assess the different B and C subclades among the 39 collected strains as well as among the 218 bacteremia O25b ST131 *E. coli* strains collected by Kallonen et al. in England between 2001 and 2012 (9), we complemented the genomes of these 257 strains with those of 21 strains representing the genetic diversity of the B and C subclades (5) and constructed a phylogenetic tree from nonrecombinant SNPs of the core genome genes using the maximum likelihood method. To infer the linear trend of the isolates of Kallonen et al. (9) from 2001 to 2012, we fitted a logistic model to the presence/absence of each subclade as a function of time (in years).

Gene deletion and complementation. Replacement by a kanamycin resistance cassette was used to inactivate genes following a strategy adapted from that of Datsenko and Wanner (49). When necessary, the kanamycin resistance-encoding gene of the cassette was eliminated, as previously described (49). Δ *fimB*::FRT and Δ *ibeA*::FRT mutants were complemented with the parental wild-type gene cloned into pSC-A-amp/kan by using a StrataClone PCR cloning kit (Agilent Technologies, Massy, France). The primers and plasmids used are listed in Tables S3 and S4.

Kinetics of early biofilm formation. The primary step (5 h of incubation) of biofilm formation was measured using the BioFilm ring test (BioFilm Control, Saint-Beauzire, France) according to the manufacturer's recommendations and as previously described (19, 28). The BFI, which had values ranging from 20 (the absence of biofilm formation) to 0 (high level of biofilm formation), is inversely proportional to the biofilm formation ability. A BFI value of 10 was chosen as the biofilm production cutoff (a BFI of ≤ 10 indicates biofilm formation, a BFI of > 10 indicates no biofilm formation).

Expression of type 1 fimbriae. Expression of type 1 fimbriae was assessed by using the yeast (*Saccharomyces cerevisiae*) cell agglutination assay, as previously described (21), and, after adaptations to highlight the early expression of type 1 fimbriae, by using 10 μ l of a pellet obtained after centrifugation ($3,000 \times g$, 10 min) of 3 ml of LB culture after 2- and 5-h incubations. This test was positive when aggregations appeared within 3 min.

Curli expression and cellulose production. Congo red assays were used to assess curli fimbria expression and cellulose production, as previously described (50). Assay positivity resulted in the red, dry, and rough (rdar) colony morphotype.

Colicin and/or phage production. Colicin and/or phage production in the strains tested was detected as previously described (51) in competition assays.

Maximum growth rate. The maximum growth rate (MGR) assay was performed as previously described (52) in LB by using an automatic spectrophotometer (Tecan Infinite F200 Pro) that measures the optical density at 600 nm in each well every 5 min over a period of 24 h. The experiment was repeated three times. Growth curves were then analyzed, and the MGR was calculated and expressed in hours^{-1} .

Mouse models. (i) Mono-infection assay in the sepsis model. The individual ability of the 39 strains and mutants to cause mouse sepsis was assessed as previously described (53). From 10 to 20 mice were used for each assay. Kaplan-Meier curves of mouse survival were performed.

(ii) Competition assays in sepsis, intestinal colonization, and UTI models. In competition assays, the relative ability of the two strains tested together (i) to cause sepsis was determined from the spleen bacterial load (53), (ii) their relative ability to colonize the intestine mouse was determined from the fecal bacterial load at days 1, 4, and 7 postinoculation (52), and (iii) their relative ability to cause an ascending unobstructed UTI was determined from the bladder and kidney bacterial load (54). From 5 to 10 mice were used for each competition assay. Competitive indexes (CI) were obtained using the following formula: $[(\log \text{number of CFU of isolate1} / \log \text{number of CFU of isolate2}) \text{ at } T_x] / [(\log \text{number of CFU of isolate1} / \log \text{number of CFU of isolate2}) \text{ at } T_0]$, with isolate1 being the first strain cited in the figure legends, isolate2 being the second one cited in the figure legends, T_x being time x , and T_0 being time zero.

Statistical analysis. The Wilcoxon rank-sum test was used to compare the average number of VF-encoding genes per clade, and Fisher's exact test was used to test the distribution of each VF-encoding gene between the clades. The individual fitness measurement was estimated by used of a mixed model with the random effect on the strain to take into account the triplicate determination. For biofilm data, a two-way repeated-measures analysis of variance, followed by Tukey's range test, when there were three groups, was performed. The association between groups (biofilm phenotypes, yeast

agglutination test, Congo red assay, and vortype assay) was assessed by Fisher's exact test. For *in vivo* competitions, a nonparametric Wilcoxon test on paired data was conducted on CI values, and *P* values were corrected for multiple comparisons by the Benjamini-Hochberg procedure (55), when necessary. Mouse survival differences were determined by the log-rank test. A significance level of <0.05 was used for all tests. All statistical analyses were carried out with R software (56).

Ethics statement. Murine protocols of sepsis, intestinal colonization, and pyelonephritis (protocol numbers APAFIS#4948- and APAFIS#4951-2016020515004032 v2, 2016021216251548 v4, and APAFIS#4950-2016021211417682 v4, respectively) were approved by the French Ministry of Research and by the Ethical Committee for Animal Experiments, CEEA-121, Comité d'éthique Paris-Nord.

Data availability. The raw sequences of the 39 isolates were deposited in GenBank under BioProject accession numbers [PRJNA320043](https://doi.org/10.6026/10.1008/PRJNA320043) and [PRJNA566165](https://doi.org/10.6026/10.1008/PRJNA566165).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 0.2 MB.

SUPPLEMENTAL FILE 2, PDF file, 0.1 MB.

SUPPLEMENTAL FILE 3, PDF file, 0.1 MB.

SUPPLEMENTAL FILE 4, PDF file, 0.2 MB.

SUPPLEMENTAL FILE 5, PDF file, 0.2 MB.

ACKNOWLEDGMENTS

We thank Christophe Beloin for his advice about biofilm exploration. We thank MÉRIL Massot, Marie Vigan, and Cedric Laouénan for their help with statistical analyses. We also thank Olivier Tenaillon and Antoine Bridier-Nahmias for genomic analysis assistance. Finally, we thank Johann Beghain, Mélanie Magnan, Françoise Chau, and Zineb Azzaoui for their technical assistance in this work.

This study was supported by a grant from the project JPI-EC-AMR 2016 with the French Agence Nationale de la Recherche (ANR) as the sponsor (grant number ANR-16-JPEC-0002-04) to M.-H.N.-C. and by the Fondation pour la Recherche Médicale (Equipe FRM 2016, grant number DEQ20161136698) to E.D. S.-C.F.-S. acknowledges the FPU program for her grant (grant FPU15/O2644) from the Secretaría General de Universidades, Spanish Ministerio de Educación, Cultura y Deporte.

REFERENCES

- Price LB, Johnson JR, Aziz M, Clabots C, Johnston B, Tchesnokova V, Nordstrom L, Billig M, Chattopadhyay S, Stegger M, Andersen PS, Pearson T, Riddell K, Rogers P, Scholes D, Kahl B, Keim P, Sokurenko EV. 2013. The epidemic of extended-spectrum- β -lactamase-producing *Escherichia coli* ST131 is driven by a single highly pathogenic subclone, H30-Rx. *mBio* 4:e00377-13. <https://doi.org/10.1128/mBio.00377-13>.
- Johnson JR, Tchesnokova V, Johnston B, Clabots C, Roberts PL, Billig M, Riddell K, Rogers P, Qin X, Butler-Wu S, Price LB, Aziz M, Nicolas-Chanoine M-H, Debroy C, Robicsek A, Hansen G, Urban C, Platell J, Trott DJ, Zhanel G, Weissman SJ, Cookson BT, Fang FC, Limaye AP, Scholes D, Chattopadhyay S, Hooper DC, Sokurenko EV. 2013. Abrupt emergence of a single dominant multidrug-resistant strain of *Escherichia coli*. *J Infect Dis* 207:919–928. <https://doi.org/10.1093/infdis/jis933>.
- Olesen B, Frimodt-Møller J, Leihof RF, Struve C, Johnston B, Hansen DS, Scheutz F, Krogfelt KA, Kuskowski MA, Clabots C, Johnson JR. 2014. Temporal trends in antimicrobial resistance, and virulence-associated traits within the *Escherichia coli* sequence type 131 clonal group and its H30 and H30-Rx subclones, 1968 to 2012. *Antimicrob Agents Chemother* 58:6886–6895. <https://doi.org/10.1128/AAC.03679-14>.
- Petty NK, Ben Zakour NL, Stanton-Cook M, Skippington E, Totsika M, Forde BM, Phan M-D, Gomes Moriel D, Peters KM, Davies M, Rogers BA, Dougan G, Rodríguez-Baño J, Pascual A, Pitout JDD, Upton M, Paterson DL, Walsh TR, Schembri MA, Beatson SA. 2014. Global dissemination of a multidrug resistant *Escherichia coli* clone. *Proc Natl Acad Sci U S A* 111:5694–5699. <https://doi.org/10.1073/pnas.1322678111>.
- Ben Zakour NL, Alsheikh-Hussain AS, Ashcroft MM, Khanh Nhu NT, Roberts LW, Stanton-Cook M, Schembri MA, Beatson SA. 2016. Sequential acquisition of virulence and fluoroquinolone resistance has shaped the evolution of *Escherichia coli* ST131. *mBio* 7:e00347-16. <https://doi.org/10.1128/mBio.00347-16>.
- Matsumura Y, Pitout JDD, Gomi R, Matsuda T, Noguchi T, Yamamoto M, Peirano G, DeVinney R, Bradford PA, Motyl MR, Tanaka M, Nagao M, Takakura S, Ichiyama S. 2016. Global *Escherichia coli* sequence type 131 clade with *bla*_{CTX-M-27} gene. *Emerg Infect Dis* 22:1900–1907. <https://doi.org/10.3201/eid2211.160519>.
- Nicolas-Chanoine M-H, Bertrand X, Madec J-Y. 2014. *Escherichia coli* ST131, an intriguing clonal group. *Clin Microbiol Rev* 27:543–574. <https://doi.org/10.1128/CMR.00125-13>.
- Morales-Barroso I, López-Cerero L, Molina J, Bellido M, Navarro MD, Serrano L, González-Galán V, Praena J, Pascual A, Rodríguez-Baño J. 2017. Bacteraemia due to non-ESBL-producing *Escherichia coli* O25b:H4 sequence type 131: insights into risk factors, clinical features and outcomes. *Int J Antimicrob Agents* 49:498–502. <https://doi.org/10.1016/j.ijantimicag.2016.12.013>.
- Kallonen T, Brodrick HJ, Harris SR, Corander J, Brown NM, Martin V, Peacock SJ, Parkhill J. 2017. Systematic longitudinal survey of invasive *Escherichia coli* in England demonstrates a stable population structure only transiently disturbed by the emergence of ST131. *Genome Res* 27:1437–1449. <https://doi.org/10.1101/gr.216606.116>.
- Blanco J, Mora A, Mamani R, López C, Blanco M, Dahbi G, Herrera A, Blanco JE, Alonso MP, García-Garrote F, Chaves F, Orellana MÁ, Martínez-Martínez L, Calvo J, Prats G, Larrosa MN, González-López JJ, López-Cerero L, Rodríguez-Baño J, Pascual A. 2011. National survey of *Escherichia coli* causing extraintestinal infections reveals the spread of drug-resistant clonal groups O25b:H4-B2-ST131, O15:H1-D-ST393 and CGA-D-ST69 with high virulence gene content in Spain. *J Antimicrob Chemother* 66:2011–2021. <https://doi.org/10.1093/jac/dkr235>.
- Ludden C, Decano AG, Jamrozny D, Pickard D, Morris D, Parkhill J, Peacock SJ, Cormican M, Downing T. 2020. Genomic surveillance of *Escherichia coli* ST131 identifies local expansion and serial replacement of subclones. *Microb Genom* 6:e000352. <https://doi.org/10.1099/mgen.0.000352>.

12. Huseby DL, Pietsch F, Brandis G, Garoff L, Tegehall A, Hughes D. 2017. Mutation supply and relative fitness shape the genotypes of ciprofloxacin-resistant *Escherichia coli*. *Mol Biol Evol* 34:1029–1039. <https://doi.org/10.1093/molbev/msx052>.
13. Johnson TJ, Danzeisen JL, Youmans B, Case K, Llop K, Munoz-Aguayo J, Flores-Figueroa C, Aziz M, Stoesser N, Sokurenko E, Price LB, Johnson JR. 2016. Separate F-type plasmids have shaped the evolution of the H30 subclone of *Escherichia coli* sequence type 131. *mSphere* 1:e00121-16. <https://doi.org/10.1128/mSphere.00121-16>.
14. McNally A, Kallonen T, Connor C, Abudahab K, Aanensen DM, Horner C, Peacock SJ, Parkhill J, Croucher NJ, Corander J. 2019. Diversification of colonization factors in a multidrug-resistant *Escherichia coli* lineage evolving under negative frequency-dependent selection. *mBio* 10:e00644-19. <https://doi.org/10.1128/mBio.00644-19>.
15. Billard-Pomares T, Clermont O, Castellanos M, Magdoud F, Royer G, Condamine B, Fouteau S, Barbe V, Roche D, Cruveiller S, Médigue C, Pognard D, Glodt J, Dion S, Rigal O, Picard B, Denamur E, Branger C. 2019. The arginine deiminase operon is responsible for a fitness trade-off in extended-spectrum- β -lactamase-producing strains of *Escherichia coli*. *Antimicrob Agents Chemother* 63:e00635-19. <https://doi.org/10.1128/AAC.00635-19>.
16. Leflon-Guibout V, Blanco J, Amaqdouf K, Mora A, Guize L, Nicolas-Chanoine M-H. 2008. Absence of CTX-M enzymes but high prevalence of clones, including clone ST131, among fecal *Escherichia coli* isolates from healthy subjects living in the area of Paris, France. *J Clin Microbiol* 46:3900–3905. <https://doi.org/10.1128/JCM.00734-08>.
17. Clermont O, Couffignal C, Blanco J, Mentré F, Picard B, Denamur E, COLVILLE and COLIBAFI Groups. 2017. Two levels of specialization in bacteraeic *Escherichia coli* strains revealed by their comparison with commensal strains. *Epidemiol Infect* 145:872–882. <https://doi.org/10.1017/S0950268816003010>.
18. Saadi S, Maas WK, Hill DF, Bergquist PL. 1987. Nucleotide sequence analysis of RepFIC, a basic replicon present in IncFI plasmids P307 and F, and its relation to the RepA replicon of IncFII plasmids. *J Bacteriol* 169:1836–1846. <https://doi.org/10.1128/jb.169.5.1836-1846.1987>.
19. Flament-Simon S-C, Duprilot M, Mayer N, García V, Alonso MP, Blanco J, Nicolas-Chanoine M-H. 2019. Association between kinetics of early biofilm formation and clonal lineage in *Escherichia coli*. *Front Microbiol* 10:1183. <https://doi.org/10.3389/fmicb.2019.01183>.
20. Fronzes R, Remaut H, Waksman G. 2008. Architectures and biogenesis of non-flagellar protein appendages in Gram-negative bacteria. *EMBO J* 27:2271–2280. <https://doi.org/10.1038/emboj.2008.155>.
21. Totsika M, Beatson SA, Sarkar S, Phan M-D, Petty NK, Bachmann N, Szubert M, Sidjabat HE, Paterson DL, Upton M, Schembri MA. 2011. Insights into a multidrug resistant *Escherichia coli* pathogen of the globally disseminated ST131 lineage: genome analysis and virulence mechanisms. *PLoS One* 6:e26578. <https://doi.org/10.1371/journal.pone.0026578>.
22. Cortes MAM, Gibon J, Chanteloup NK, Moulin-Schouleir M, Gilot P, Germon P. 2008. Inactivation of *ibeA* and *ibeT* results in decreased expression of type 1 fimbriae in extraintestinal pathogenic *Escherichia coli* strain BEN2908. *Infect Immun* 76:4129–4136. <https://doi.org/10.1128/IAI.00334-08>.
23. Johnson JR, Clermont O, Menard M, Kuskowski MA, Picard B, Denamur E. 2006. Experimental mouse lethality of *Escherichia coli* isolates, in relation to accessory traits, phylogenetic group, and ecological source. *J Infect Dis* 194:1141–1150. <https://doi.org/10.1086/507305>.
24. Picard B, García JS, Gouriou S, Duriez P, Brahimi N, Bingen E, Elion J, Denamur E. 1999. The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection. *Infect Immun* 67:546–553. <https://doi.org/10.1128/IAI.67.2.546-553.1999>.
25. Turret J, Diard M, Garry L, Matic I, Denamur E. 2010. Effects of single and multiple pathogenicity island deletions on uropathogenic *Escherichia coli* strain 536 intrinsic extra-intestinal virulence. *Int J Med Microbiol* 300:435–439. <https://doi.org/10.1016/j.ijmm.2010.04.013>.
26. Dahbi G, Mora A, Mamani R, López C, Alonso MP, Marzoa J, Blanco M, Herrera A, Viso S, García-Garrote F, Tcheshnokova V, Billig M, de la Cruz F, de Toro M, González-López JJ, Prats G, Chaves F, Martínez-Martínez L, López-Cerezo L, Denamur E, Blanco J. 2014. Molecular epidemiology and virulence of *Escherichia coli* O16:H5-ST131: comparison with H30 and H30-Rx subclones of O25b:H4-ST131. *Int J Med Microbiol* 304:1247–1257. <https://doi.org/10.1016/j.ijmm.2014.10.002>.
27. Lavigne J-P, Vergunst AC, Goret L, Sotto A, Combescuré C, Blanco J, O'Callaghan D, Nicolas-Chanoine M-H. 2012. Virulence potential and genomic mapping of the worldwide clone *Escherichia coli* ST131. *PLoS One* 7:e34294. <https://doi.org/10.1371/journal.pone.0034294>.
28. Nicolas-Chanoine M-H, Petitjean M, Mora A, Mayer N, Lavigne J-P, Boulet O, Leflon-Guibout V, Blanco J, Hocquet D. 2017. The ST131 *Escherichia coli* H22 subclone from human intestinal microbiota: comparison of genomic and phenotypic traits with those of the globally successful H30 subclone. *BMC Microbiol* 17:71. <https://doi.org/10.1186/s12866-017-0984-8>.
29. Le Gall T, Clermont O, Gouriou S, Picard B, Nassif X, Denamur E, Tenaillon O. 2007. Extraintestinal virulence is a coincidental by-product of commensalism in B2 phylogenetic group *Escherichia coli* strains. *Mol Biol Evol* 24:2373–2384. <https://doi.org/10.1093/molbev/msm172>.
30. Gunther NW, Snyder JA, Lockatell V, Blomfield I, Johnson DE, Mobley HLT. 2002. Assessment of virulence of uropathogenic *Escherichia coli* type 1 fimbrial mutants in which the invertible element is phase-locked on or off. *Infect Immun* 70:3344–3354. <https://doi.org/10.1128/IAI.70.7.3344-3354.2002>.
31. Schwan WR, Ding H. 2017. Temporal regulation of *fim* genes in uropathogenic *Escherichia coli* during infection of the murine urinary tract. *J Pathog* 2017:8694356. <https://doi.org/10.1155/2017/8694356>.
32. Brisse S, Diancourt L, Lauouénan C, Vigan M, Caro V, Arlet G, Drieux L, Leflon-Guibout V, Mentré F, Jarlier V, Nicolas-Chanoine M-H, Coli β Study Group. 2012. Phylogenetic distribution of CTX-M- and non-extended-spectrum- β -lactamase-producing *Escherichia coli* isolates: group B2 isolates, except clone ST131, rarely produce CTX-M enzymes. *J Clin Microbiol* 50:2974–2981. <https://doi.org/10.1128/JCM.00919-12>.
33. Johnson JR, Porter SB, Zhanel G, Kuskowski MA, Denamur E. 2012. Virulence of *Escherichia coli* clinical isolates in a murine sepsis model in relation to sequence type ST131 status, fluoroquinolone resistance, and virulence genotype. *Infect Immun* 80:1554–1562. <https://doi.org/10.1128/IAI.06388-11>.
34. Mora A, Dahbi G, López C, Mamani R, Marzoa J, Dion S, Picard B, Blanco M, Alonso MP, Denamur E, Blanco J. 2014. Virulence patterns in a murine sepsis model of ST131 *Escherichia coli* clinical isolates belonging to serotypes O25b:H4 and O16:H5 are associated to specific virotypes. *PLoS One* 9:e87025. <https://doi.org/10.1371/journal.pone.0087025>.
35. Merino I, Porter SB, Johnston B, Clabots C, Thuras P, Ruiz-Garbjosa P, Cantón R, Johnson JR. 2020. Molecularly defined extraintestinal pathogenic *Escherichia coli* status predicts virulence in a murine sepsis model better than does virotype, individual virulence genes, or clonal subset among *E. coli* ST131 isolates. *Virulence* 11:327–336. <https://doi.org/10.1080/21505594.2020.1747799>.
36. Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. 2016. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biol* 17:238. <https://doi.org/10.1186/s13059-016-1108-8>.
37. Galardini M, Clermont O, Baron A, Busby B, Dion S, Schubert S, Beltrao P, Denamur E. 2020. Major role of iron uptake systems in the intrinsic extra-intestinal virulence of the genus *Escherichia* revealed by a genome-wide association study. *bioRxiv* 712034. <https://doi.org/10.1101/712034>.
38. Basra P, Alsaadi A, Bernal-Astrain G, O'Sullivan ML, Hazlett B, Clarke LM, Schoenrock A, Pitre S, Wong A. 2018. Fitness tradeoffs of antibiotic resistance in extraintestinal pathogenic *Escherichia coli*. *Genome Biol Evol* 10:667–679. <https://doi.org/10.1093/gbe/evy030>.
39. Anderson RM, May RM. 1982. Coevolution of hosts and parasites. *Parasitology* 85(Pt 2):411–426. <https://doi.org/10.1017/s0031182000055360>.
40. Diard M, Hardt W-D. 2017. Evolution of bacterial virulence. *FEMS Microbiol Rev* 41:679–697. <https://doi.org/10.1093/femsre/fux023>.
41. Liu CM, Stegger M, Aziz M, Johnson TJ, Waits K, Nordstrom L, Gauld L, Weaver B, Rolland D, Statham S, Horwinski J, Sariya S, Davis GS, Sokurenko E, Keim P, Johnson JR, Price LB. 2018. *Escherichia coli* ST131-H22 as a foodborne uropathogen. *mBio* 9:e00470-18. <https://doi.org/10.1128/mBio.00470-18>.
42. Salvador E, Wagenlehner F, Köhler C-D, Mellmann A, Hacker J, Svanborg C, Dobrindt U. 2012. Comparison of asymptomatic bacteriuria *Escherichia coli* isolates from healthy individuals versus those from hospital patients shows that long-term bladder colonization selects for attenuated virulence phenotypes. *Infect Immun* 80:668–678. <https://doi.org/10.1128/IAI.06191-11>.
43. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, Møller Aarestrup F, Hasman H. 2014. *In silico* detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother* 58:3895–3903. <https://doi.org/10.1128/AAC.02412-14>.

44. Seemann T. 2018. Abricate. <https://github.com/tseemann/abricate>.
45. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV. 2012. Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* 67:2640–2644. <https://doi.org/10.1093/jac/dks261>.
46. Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS, Nielsen EM, Aarestrup FM. 2014. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *J Clin Microbiol* 52:1501–1510. <https://doi.org/10.1128/JCM.03617-13>.
47. Chen L, Zheng D, Liu B, Yang J, Jin Q. 2016. VFDB 2016: hierarchical and refined dataset for big data analysis—10 years on. *Nucleic Acids Res* 44:D694–D697. <https://doi.org/10.1093/nar/gkv1239>.
48. Vallenet D, Engelen S, Mornico D, Cruveiller S, Fleury L, Lajus A, Rouy Z, Roche D, Salvignol G, Scarpelli C, Médigue C. 2009. MicroScope: a platform for microbial genome annotation and comparative genomics. *Database (Oxford)* 2009:bap021. <https://doi.org/10.1093/database/bap021>.
49. Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* 97:6640–6645. <https://doi.org/10.1073/pnas.120163297>.
50. Dudin O, Geiselmann J, Ogasawara H, Ishihama A, Lacour S. 2014. Repression of flagellar genes in exponential phase by CsgD and CpxR, two crucial modulators of *Escherichia coli* biofilm formation. *J Bacteriol* 196:707–715. <https://doi.org/10.1128/JB.00938-13>.
51. Smati M, Clermont O, Bleibtreu A, Fourreau F, David A, Daubié A-S, Hignard C, Loison O, Picard B, Denamur E. 2015. Quantitative analysis of commensal *Escherichia coli* populations reveals host-specific enterotypes at the intra-species level. *Microbiologyopen* 4:604–615. <https://doi.org/10.1002/mbo3.266>.
52. Vimont S, Boyd A, Bleibtreu A, Bens M, Goujon J-M, Garry L, Clermont O, Denamur E, Arlet G, Vandewalle A. 2012. The CTX-M-15-producing *Escherichia coli* clone O25b:H4-ST131 has high intestine colonization and urinary tract infection abilities. *PLoS One* 7:e46547. <https://doi.org/10.1371/journal.pone.0046547>.
53. Smati M, Magistro G, Adiba S, Wieser A, Picard B, Schubert S, Denamur E. 2017. Strain-specific impact of the high-pathogenicity island on virulence in extra-intestinal pathogenic *Escherichia coli*. *Int J Med Microbiol* 307:44–56. <https://doi.org/10.1016/j.ijmm.2016.11.004>.
54. Labat F, Pradillon O, Garry L, Peuchmaur M, Fantin B, Denamur E. 2005. Mutator phenotype confers advantage in *Escherichia coli* chronic urinary tract infection pathogenesis. *FEMS Immunol Med Microbiol* 44:317–321. <https://doi.org/10.1016/j.femsim.2005.01.003>.
55. Holm S. 1979. A simple sequentially rejective multiple test procedure. *Scand J Stat* 6:65–70.
56. R Development Core Team. 2014. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.